scientific reports

Check for updates

OPEN

Multilingual sentiment analysis in restaurant reviews using aspect focused learning

Arifur Rahman[1], Md. Azam Khan[1], Kanchon Kumar Bishnu[2], Uland Rozario[3], Adit Ishraq[4], M. F. Mridha[5] & Zeyar Aung[6✉]

Cross-cultural sentiment analysis in restaurant reviews presents unique challenges due to linguistic and cultural differences across regions. The purpose of this study is to develop a culturally adaptive sentiment analysis model that improves sentiment detection across multilingual restaurant reviews. This paper proposes XLM-RSA, a novel multilingual model based on XLM-RoBERTa with Aspect-Focused Attention, tailored for enhanced sentiment analysis across diverse cultural contexts. We evaluated XLM-RSA on three benchmark datasets: 10,000 Restaurant Reviews, Restaurant Reviews, and European Restaurant Reviews, achieving state-of-the-art performance across all datasets. XLM-RSA attained an accuracy of 91.9% on the Restaurant Reviews dataset, surpassing traditional models such as BERT (87.8%) and RoBERTa (88.5%). In addition to sentiment classification, we introduce an aspect-based attention mechanism to capture sentiment variations specific to key aspects like food, service, and ambiance, yielding aspect-level accuracy improvements. Furthermore, XLM-RSA demonstrated strong performance in detecting cultural sentiment shifts, with an accuracy of 85.4% on the European Restaurant Reviews dataset, showcasing its robustness to diverse linguistic and cultural expressions. An ablation study highlighted the significance of the Aspect-Focused Attention, where XLM-RSA with this enhancement achieved an F1-score of 91.5%, compared to 89.1% with a simple attention mechanism. These results affirm XLM-RSA's capacity for effective cross-cultural sentiment analysis, paving the way for more accurate sentiment-driven insights in globally distributed customer feedback.

Keywords Cross-cultural sentiment analysis, XLM-RoBERTa, Aspect-focused attention, Restaurant reviews, Natural language processing, Sentiment classification, Deep learning

Sentiment analysis in online customer reviews is a pivotal tool for understanding public perception and consumer satisfaction. It has widespread application in fields such as marketing, service improvement, and customer relationship management[1]. With the increasing accessibility of online reviews from different geographic regions, analyzing sentiments across multiple cultures has become an essential but complex task[2]. Variations in language, cultural expressions, and review structures complicate traditional sentiment analysis models, which are generally trained on monolingual or culturally homogeneous datasets[3]. This study addresses these challenges by developing and implementing a multilingual model, XLM-RSA, for cross-cultural sentiment analysis in restaurant reviews.

Cross-cultural sentiment analysis involves interpreting sentiment across diverse linguistic and cultural backgrounds[4]. Traditional sentiment analysis models often fail to capture subtle semantic and syntactic variations that are unique to different languages and cultural contexts[5]. For instance, sentiments expressed in Asian restaurant reviews may emphasize humility and subtlety, whereas Western reviews often employ direct language[6,7]. This divergence in expression can lead to significant misclassification in sentiment analysis tasks if a model lacks adaptation to such nuances.

Multilingual language models, such as mBERT[8] and XLM-RoBERTa[9], have introduced a new paradigm in natural language processing by enabling sentiment analysis across languages without requiring extensive labeled data in each language. However, while effective, these models are often limited in handling specific cultural nuances in sentiment. Moreover, previous approaches to cross-cultural sentiment analysis typically fail to capture aspect-level sentiment, where individual elements of a service (e.g., food, ambiance, service) contribute

[1]School of Business, International American University, Los Angeles, CA, USA. [2]Department of Computer Science, California State University, Los Angeles, CA, USA. [3]Department of CSE, Bangladesh University of Business and Technology, Dhaka 1216, Bangladesh. [4]Stockholm University, Frescativägen, 114 19 Stockholm, Sweden. [5]Department of Computer Science, American International University-Bangladesh, Dhaka 1229, Bangladesh. [6]Department of Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates. ✉email: zeyar.aung@ku.ac.ae

distinctly to overall customer satisfaction[10]. Aspect-based sentiment analysis (ABSA) has proven useful in detailed sentiment breakdowns but lacks an integrated approach for cultural sentiment shifts[11].

Cultural sensitivity is essential for the reliable sentiment analysis of online reviews. Variations in cultural norms and communication styles affect how sentiments are expressed, with potential differences in tone, directness, and emphasis. For instance, reviews from cultures where indirect or polite criticism is customary might be misinterpreted by traditional sentiment models, resulting in inaccurate sentiment classification[12]. Addressing this variability is essential for obtaining precise, context-sensitive insights, especially in hospitality where customer experiences and expectations are deeply culture-bound.

Another key challenge is the limited availability of annotated multilingual data that adequately represent diverse linguistic and cultural backgrounds. Although multilingual models such as mBERT and XLM-RoBERTa have enhanced cross-lingual adaptability, they still require fine-tuning with culturally representative data to handle specific nuances[13]. This lack of culturally labeled data frequently limits the model's cross-regional generalization, underscoring the need for more flexible and resource-efficient models that can use sparse multilingual data while still achieving high accuracy. Aspect-based sentiment analysis offers a fine-grained perspective by focusing on specific aspects within a review, such as food quality, ambiance, or service in the context of restaurant reviews. However, traditional ABSA models lack the cross-cultural adaptability to capture how sentiments vary across cultures. Our study addresses this limitation by introducing an aspect-focused attention mechanism within a multilingual framework, allowing for more detailed and culturally aware sentiment breakdowns that align with diverse customer expectations.

To further improve sentiment analysis in a cross-cultural context, it is essential to consider the challenges posed by implicit sentiment expressions. Cultural differences often affect how sentiment is implied rather than explicitly stated, which can be especially prevalent in indirect communication styles. Existing models may struggle with these nuances because they tend to rely on explicit sentiment cues. Thus, there is a need for models that can better infer sentiment from subtle or context-dependent language cues that are influenced by cultural norms and expressions. In addition, cross-cultural sentiment analysis of restaurant reviews presents unique challenges because customer priorities can vary significantly by region. For example, while Western reviews may prioritize factors such as customer service speed and convenience, Eastern reviews may emphasize qualities like authenticity and traditional ambiance. By focusing on these regional preferences and customer expectations, sentiment analysis models can offer more accurate insights that cater to local values. This requires integrating culturally sensitive feature extraction methods, which can distinguish these varying priorities in sentiment classification.

To bridge these gaps, this paper introduces XLM-RSA, a model that fine-tunes XLM-RoBERTa with an added aspect-focused Attention mechanism. This enhancement allows the model to focus on sentiment-rich keywords related to specific aspects within a review, thereby improving sentiment classification accuracy in a multilingual and culturally diverse context. By combining cross-lingual embeddings with aspect-level attention, XLM-RSA achieves improved sentiment recognition across different regions while accurately identifying sentiment nuances that reflect cultural variability.

This work contributes to the literature in several ways:

- Novel Architecture Design: We propose XLM-RSA, a sentiment analysis model that integrates aspect-based attention with XLM-RoBERTa, tailored for multilingual sentiment analysis across cultures.
- Enhanced Cross-Cultural Sentiment Recognition: XLM-RSA includes a sentiment-focused loss function and a cultural sentiment shift detection mechanism, achieving robust performance across datasets with varied linguistic and cultural backgrounds.
- Aspect-Level Sentiment Detection: The model's aspect-focused attention layer enables detailed sentiment classification for key aspects such as food, service, and ambiance, addressing the need for finer granularity in cross-cultural reviews.
- State-of-the-Art Performance: Experimental results demonstrate that XLM-RSA outperforms traditional models like mBERT and XLM-RoBERTa alone, achieving high accuracy, precision, and F1 scores across three datasets–10,000 Restaurant Reviews, Restaurant Reviews, and European Restaurant Reviews.
- Comprehensive Ablation Study: An extensive ablation study illustrates the effectiveness of aspect-focused attention, with a significant increase in sentiment classification accuracy and aspect-specific predictions compared to simpler attention mechanisms.

The remainder of this paper is structured as follows: Section "Related work" reviews related work in multilingual and cross-cultural sentiment analysis. Section "Methodology" details the proposed XLM-RSA architecture, while Section "Results" discusses the datasets and experimental setup. Section Results presents the results, followed by an ablation study in Section "Discussion". Finally, conclusions and future work are provided in Section "Conclusions".

Related work
The development of robust sentiment analysis models for cross-cultural and multilingual data has become an active area of research in recent years. This section reviews the advances in multilingual sentiment analysis, cross-cultural sentiment analysis, and aspect-based sentiment analysis, laying the foundation for the proposed XLM-RSA model.

Aspect-Based Sentiment Analysis has evolved into multiple subtasks beyond basic classification. The most common is Aspect-Based Sentiment Classification (ASC), which identifies the sentiment polarity associated with predefined aspects (e.g., "food" or "service")[14]. More advanced subtasks include Aspect-Based Sentiment Triplet Extraction (ASTE), which extracts aspect terms, opinion terms, and sentiment polarities jointly[15], and Aspect-Based Sentiment Quadruple Extraction (ASQP), which extends this by including the opinion holder

or target[16]. Additionally, recent work in Dimensional ABSA (dimABSA) focuses on predicting continuous sentiment scores (e.g., valence) at the aspect level, offering a finer-grained emotional understanding[17]. These developments highlight the richness of ABSA as a task.

Multilingual sentiment analysis seeks to analyze sentiments across multiple languages, often without the need for language-specific training data. Pre-trained multilingual models, such as mBERT[18] and XLM-RoBERTa[19], have demonstrated significant improvements in handling diverse languages within a single framework. mBERT, as a multilingual extension of BERT, applies masked language modeling across multiple languages, achieving strong performance on cross-lingual tasks[20]. Similarly, XLM-RoBERTa extends RoBERTa's architecture to over 100 languages, achieving state-of-the-art results in various multilingual NLP tasks[21]. However, these models face challenges in addressing cultural sentiment nuances that emerge due to language and cultural diversity in datasets. Research has shown that purely multilingual models may overlook cultural context, which is essential for interpreting sentiment in nuanced settings[22]. Moreover, research is expanding to detect sentiment and mental health indicators like depression across diverse linguistic contexts. For instance, screening for depression using NLP has gained traction, as seen in recent studies[23], and multilingual depression detection approaches for social media across multiple Indian languages highlight the challenges and advances in sentiment interpretation across culturally diverse linguistic regions. Thus, while multilingual transformers are effective, there remains a gap in their ability to model sentiment shifts across cultural boundaries.

Cross-cultural sentiment analysis goes beyond language differences, addressing how sentiment expression varies according to cultural norms, expressions, and regional linguistic features. Prior studies indicate that sentiment expression is influenced by cultural norms, such as politeness in Japanese reviews and directness in American reviews[24,25]. To capture such variations, some approaches leverage transfer learning and domain adaptation techniques to align sentiment expressions across cultures[26]. Cross-lingual embeddings have also been used to bridge cultural differences by mapping semantically similar words across languages into shared vector spaces[27]. Despite these efforts, a critical limitation remains: traditional models often fail to capture aspect-specific sentiment nuances within each culture. Nakayama and Wan (2019) explored this issue by examining how sentiment balance at the aspect level affects perceived helpfulness in online reviews of subjective goods, showing that Japanese and Western consumers value different aspects, such as cost savings versus service quality, respectively[28]. Similarly, Gao et al. (2021) investigated the influence of extreme sentiments and emotions on crowdfunding performance, revealing that Western and Eastern cultural differences significantly impact how sentiment and specific emotions affect crowdfunding success and backer engagement[29]. This study addresses this gap by integrating aspect-based sentiment analysis with a culturally adaptive model, providing a more comprehensive approach to cross-cultural sentiment detection.

Aspect-based sentiment Analysis has been widely adopted to provide granular sentiment insights by analyzing specific aspects of an entity, such as food, ambiance, and service in restaurant reviews[30]. Traditional ABSA models, such as those based on Conditional Random Fields (CRFs) and SVMs, have proven effective for single-language, aspect-specific sentiment detection[31]. With the advent of deep learning, advanced ABSA models using attention mechanisms, such as the Interactive Attention Network (IAN)[32], have enabled more precise aspect sentiment extraction by focusing on aspect-related words in each sentence[33]. However, in a multilingual and cross-cultural context, ABSA models are rarely optimized for cultural sentiment shifts, limiting their applicability in diverse settings. For instance, Jiang et al. (2019) introduced a Multi-Aspect Multi-Sentiment (MAMS) dataset to address the limitations in existing ABSA datasets, where sentences often contain only one or multiple aspects with the same sentiment polarity, thus pushing forward research in nuanced aspect-level sentiment analysis[34]. Recent works have suggested using multi-head attention to enhance focus on sentiment-rich phrases across different languages[35], but few models leverage both aspect-level insights and cultural sentiment adaptation.

Recent advancements in large language models (LLMs) have opened new directions for aspect-based sentiment analysis. Instruction-tuned LLMs, such as those incorporating retrieval-based example ranking, have demonstrated strong performance on ABSA tasks by aligning model outputs with task-specific instructions and semantically relevant exemplars[36]. These methods enable flexible adaptation to various ABSA subtasks with minimal supervision. Additionally, contrastive learning approaches, such as SoftMCL pre-training, have been proposed to enhance the discriminative capacity of aspect representations through soft clustering and mutual contrast objectives[37]. Such techniques promote robust feature learning and have shown improved generalization, particularly in low-resource and multilingual scenarios. Despite these promising developments, their application to cross-cultural sentiment analysis remains limited. Furthermore, in the domain of Dimensional ABSA (dimABSA), the shared task overview by Zhang et al.[38] provides a comprehensive benchmark for evaluating models on continuous sentiment prediction at the aspect level. These recent contributions underscore the growing importance of flexible and semantically rich representations in sentiment modeling across linguistic and cultural domains.

Attention mechanisms have become a core component of modern NLP, particularly for sentiment analysis. Multi-head attention allows models to focus on different parts of a text simultaneously, which is particularly useful in capturing nuanced expressions in sentiment[39]. Transformer-based models, such as BERT and RoBERTa, have demonstrated that attention layers can significantly enhance performance in sentiment analysis by focusing on sentiment-bearing keywords[40]. Several works have extend this concept to ABSA, where attention is specifically directed to aspect-related words[41]. Zulqarnain et al.[42] introduced an efficient two-state GRU (TS-GRU) model with feature attention mechanisms, which enhances sentiment polarity detection by capturing the sequential relationships in text and leveraging feature-based attention for improved accuracy on benchmark datasets. Lian et al.[43] proposed a conversational emotion analysis framework that utilizes attention mechanisms to fuse acoustic and lexical features while also incorporating speaker embeddings to capture speaker-specific interaction patterns during dialogues, showing a significant performance improvement over state-of-the-art methods. Despite these advances, limited research has explored attention mechanisms

in cross-cultural sentiment settings. Our work builds on these advances by incorporating an Aspect-Focused Attention layer within a multilingual transformer, enabling finer attention to aspect-related sentiment in culturally diverse contexts.

Despite advancements in multilingual models, cross-cultural sentiment analysis, and aspect-based sentiment extraction, existing models struggle with comprehensive, culturally aware sentiment analysis. Many multilingual models lack aspect-level granularity, while traditional ABSA approaches do not generalize well across languages or cultural contexts[21]. Our proposed model, XLM-RSA, addresses these limitations by leveraging XLM-RoBERTa's multilingual capabilities[44] and incorporating Aspect-Focused Attention to capture both aspect-specific and culturally nuanced sentiment. By bridging these gaps, XLM-RSA aims to set a new standard for cross-cultural sentiment analysis, offering a robust solution for understanding sentiment across linguistic and cultural boundaries[40].

Methodology

The proposed XLM-RSA model is designed for cross-cultural sentiment analysis with a focus on aspect-based attention within a multilingual framework. The architecture is based on XLM-RoBERTa with modifications to enhance sentiment classification across languages and cultural nuances. This section details each component of the model, supported by mathematical formulations.

Data preprocessing

Effective data preprocessing is essential to prepare the multilingual and cross-cultural restaurant review datasets for sentiment analysis. In this section, we describe the preprocessing techniques applied to the raw text data, including tokenization, embedding generation, noise reduction, translation, and aspect extraction. The goal is to standardize text inputs across languages and reduce variability in sentiment representation.

Tokenization and normalization
Given that our data contains text in multiple languages, we employ the XLM-RoBERTa tokenizer to tokenize each sentence, preserving the semantic nuances across languages. For a sentence S consisting of n words, we define the tokenization function as:

$$T(S) = \{t_1, t_2, \ldots, t_m\} \tag{1}$$

where T is the tokenizer, and $m \geq n$ due to subword tokenization in XLM-RoBERTa. Each word is represented by one or more subwords to handle out-of-vocabulary (OOV) tokens effectively.

Multilingual embedding generation
Each token t_i is then converted into an embedding vector $\mathbf{e}_i$ through XLM-RoBERTa's embedding layer, capturing cross-lingual semantics. For a tokenized sequence $T(S) = \{t_1, t_2, \ldots, t_m\}$, the embeddings $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m\}$ are computed as:

$$\mathbf{e}_i = \text{Embedding}(t_i) \quad \forall i \in \{1, 2, \ldots, m\} \tag{2}$$

The resulting embeddings are aggregated to produce a sentence representation $\mathbf{E}(S)$ by summing or averaging token embeddings:

$$\mathbf{E}(S) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{e}_i \tag{3}$$

Noise reduction
Noise reduction involves removing irrelevant symbols, special characters, and extra whitespaces, which may vary across languages. Let S_{clean} represent the cleaned sentence:

$$S_{\text{clean}} = \text{Clean}(S) \tag{4}$$

where the cleaning function removes non-informative tokens without altering meaningful content.

Translation and back-translation for data augmentation
To augment the dataset and enhance the model's adaptability to linguistic variations, we applied translation and back-translation techniques during training, particularly for non-English sentences in the European Restaurant Reviews dataset. In our implementation, we selected English as the auxiliary language (L') for translation of non-English sentences such as French and German. Each sentence was first translated to English, then back to the original language to introduce variation while preserving semantic meaning. To augment the dataset and enhance the model's adaptability to linguistic variations, we apply translation and back-translation techniques. For a sentence S in language L, we translate it to an auxiliary language L' and back to L, generating a syntactically varied but semantically equivalent sentence S':

$$S' = \text{BackTranslate}(\text{Translate}(S, L \rightarrow L), L' \rightarrow L) \tag{5}$$

Aspect-based extraction

Aspect-based sentiment analysis requires the identification of key aspects (e.g., food, service, ambiance) within each sentence. Let A denote the set of aspects, where $A = \{a_1, a_2, \ldots, a_k\}$. For each sentence S, we extract relevant aspect terms $\{a_{i_1}, a_{i_2}, \ldots, a_{i_l}\} \subseteq A$ based on a rule-based or machine learning-based approach:

$$\text{Aspects}(S) = \{a_{i_1}, a_{i_2}, \ldots, a_{i_l}\} \quad \text{where } a_{i_j} \in A \tag{6}$$

The aspect terms were highlighted using attention weights in the model to focus on the sentiment-bearing words associated with each aspect.

Sentiment annotation and cross-lingual alignment

The next step involves aligning sentiment labels across languages to maintain consistency in sentiment interpretation. For each sentence S, a sentiment label $y \in \{-1, 0, 1\}$ (representing negative, neutral, and positive sentiment) is assigned based on labeled data. We apply cross-lingual sentiment mapping to ensure that sentiment labels align semantically across cultures:

$$y = \text{SentimentMap}(S, L) \tag{7}$$

where SentimentMap is a function that aligns sentiment scores across languages L.

Final input representation

The final input representation for each sentence S consists of a combination of token embeddings, aspect terms, and a cross-lingual sentiment label:

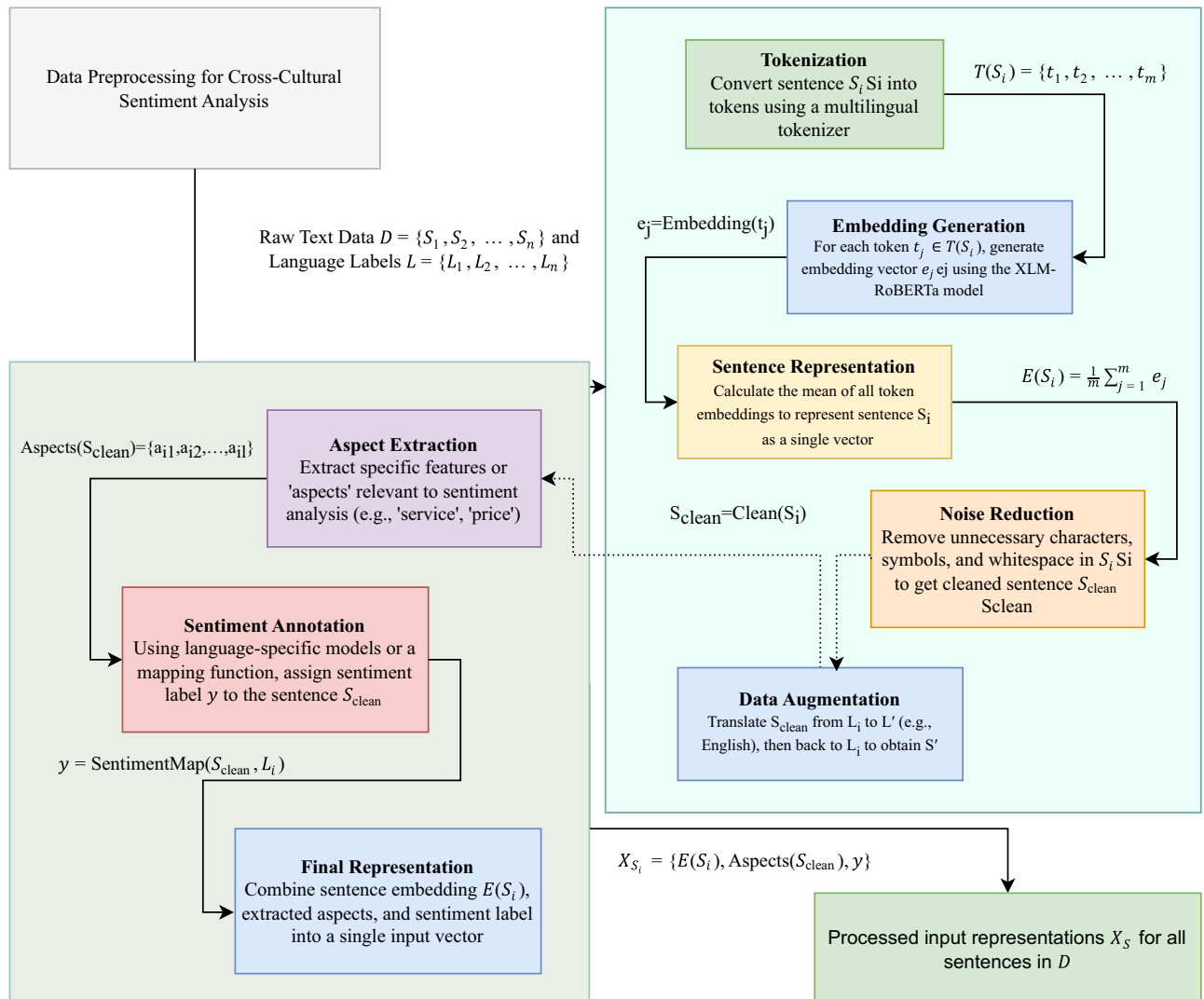$$\mathbf{X}_S = \{\mathbf{E}(S), \text{Aspects}(S), y\} \tag{8}$$

Fig. 1. Diagram of data preprocessing algorithm for cross-cultural sentiment analysis.

where $\mathbf{X}_S$ is the input to the XLM-RSA model, encoding both linguistic and cultural nuances for effective cross-cultural sentiment analysis.

Data preprocessing algorithm
The data preprocessing algorithm in Algorithm 1 outlines a structured approach for preparing multilingual restaurant review data for cross-cultural sentiment analysis. Figure 1 provides a visual representation of the data preprocessing steps, from tokenization and embedding generation to noise reduction, aspect extraction, and sentiment annotation, for preparing multilingual data for sentiment analysis. Each sentence S_i in the dataset D is first tokenized using the XLM-RoBERTa tokenizer to produce a sequence of subwords $T(S_i)$, followed by the generation of multilingual embeddings $\mathbf{e}_j$ for each token t_j within $T(S_i)$. These embeddings are aggregated into a sentence representation $\mathbf{E}(S_i)$, capturing the semantic nuances of each review across different languages. Afterward, noise reduction is applied to remove non-informative characters, and data augmentation is performed using translation and back-translation techniques to enhance linguistic variability. Aspect extraction identifies key sentiment-bearing aspects in each sentence, while sentiment annotation aligns labels across languages using cross-lingual sentiment mapping. Finally, the processed input representation $\mathbf{X}_{S_i}$ is constructed for each sentence S_i, encapsulating the embeddings, aspects, and sentiment label, thereby preparing the data for effective cross-cultural sentiment modeling.

Require: Raw text data $D = \{S_1, S_2, \ldots, S_n\}$ with language labels $L = \{L_1, L_2, \ldots, L_n\}$
Ensure: Processed input representations $\mathbf{X}_S$ for each sentence S
    for each sentence $S_i \in D$ **do**
2:    **Tokenize** S_i using XLM-RoBERTa tokenizer:

$$T(S_i) = \{t_1, t_2, \ldots, t_m\}$$

    Generate multilingual embeddings for each token $t_j \in T(S_i)$:

$$\mathbf{e}_j = \text{Embedding}(t_j)$$

4:    Aggregate embeddings to form sentence representation $\mathbf{E}(S_i)$:

$$\mathbf{E}(S_i) = \frac{1}{m}\sum_{j=1}^{m}\mathbf{e}_j$$

    Noise Reduction: Clean non-informative characters and symbols in S_i:

$$S_{\text{clean}} = \text{Clean}(S_i)$$

6:    **Data Augmentation**: Apply translation and back-translation:

$$S' = \text{BackTranslate}(\text{Translate}(S_{\text{clean}}, L_i \rightarrow L'), L' \rightarrow L_i)$$

    Aspect Extraction: Identify relevant aspects A in S_{clean}:

$$\text{Aspects}(S_{\text{clean}}) = \{a_{i_1}, a_{i_2}, \ldots, a_{i_l}\}$$

8:    **Sentiment Annotation**: Assign sentiment label y using cross-lingual alignment:

$$y = \text{SentimentMap}(S_{\text{clean}}, L_i)$$

    Form the final input representation $\mathbf{X}_{S_i}$:

$$\mathbf{X}_{S_i} = \{\mathbf{E}(S_i), \text{Aspects}(S_{\text{clean}}), y\}$$

10: **end for**
    return Processed input representations $\mathbf{X}_S$ for all sentences $S \in D$

Algorithm 1. Data preprocessing for cross-cultural sentiment analysis.

Model implementation

Input processing and embedding

Each input sentence S in a dataset D is tokenized and embedded using the XLM-RoBERTa embedding layer. For a given sentence $S = \{w_1, w_2, \ldots, w_n\}$ with n words, we apply tokenization to create a sequence of tokens $T(S) = \{t_1, t_2, \ldots, t_m\}$, where $m \geq n$ due to subword tokenization. Each token t_i is mapped to an embedding vector $\mathbf{e}_i$:

$$\mathbf{e}_i = \text{XLM-RoBERTa\_Embedding}(t_i) \quad \forall i \in \{1, \ldots, m\}. \tag{9}$$

The output embeddings $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m\}$ are passed through the XLM-RoBERTa transformer layers, resulting in contextualized embeddings $\mathbf{h}_i$ for each token:

$$\mathbf{h}_i = \text{XLM-RoBERTa}(\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m\}). \tag{10}$$

The final sentence representation, $\mathbf{H}(S)$, is derived by concatenating the contextual embeddings for all tokens:

$$\mathbf{H}(S) = [\mathbf{h}_1; \mathbf{h}_2; \ldots; \mathbf{h}_m]. \tag{11}$$

Aspect-focused attention mechanism

To capture aspect-specific sentiments, we introduce an aspect-focused attention layer that learns attention weights for specific aspects (e.g., food, service, ambiance). Let $A = \{a_1, a_2, \ldots, a_k\}$ represent the set of predefined aspects. For each aspect $a_j \in A$, we calculate an attention score α_{ij} for each token embedding $\mathbf{h}_i$ based on its relevance to the aspect a_j. The attention score is given by:

$$\alpha_{ij} = \frac{\exp(\mathbf{w}_j^\top \mathbf{h}_i)}{\sum_{i=1}^{m} \exp(\mathbf{w}_j^\top \mathbf{h}_i)} \tag{12}$$

where $\mathbf{w}_j$ is a learned weight vector specific to aspect a_j. The weighted aspect embedding $\mathbf{h}_{a_j}$ for aspect a_j is computed by summing over all token embeddings, weighted by their attention scores:

$$\mathbf{h}_{a_j} = \sum_{i=1}^{m} \alpha_{ij} \mathbf{h}_i. \tag{13}$$

The aspect embeddings for all aspects are then concatenated to form the aspect-focused sentence representation $\mathbf{H}_A(S)$:

$$\mathbf{H}_A(S) = [\mathbf{h}_{a_1}; \mathbf{h}_{a_2}; \ldots; \mathbf{h}_{a_k}]. \tag{14}$$

Sentiment classification head

The sentiment classification layer uses the aspect-focused representation $\mathbf{H}_A(S)$ to predict sentiment. We apply a fully connected layer followed by a softmax activation to output sentiment probabilities. Let $\mathbf{W}_{\text{sent}}$ and $\mathbf{b}_{\text{sent}}$ be the weights and bias for the sentiment classification layer:

$$\mathbf{z} = \mathbf{W}_{\text{sent}} \mathbf{H}_A(S) + \mathbf{b}_{\text{sent}}. \tag{15}$$

The predicted sentiment probabilities $\mathbf{p}$ are computed using the softmax function:

$$\mathbf{p} = \text{softmax}(\mathbf{z}). \tag{16}$$

The model's predicted sentiment label $\hat{y}$ for each sentence S is then:

$$\hat{y} = \arg\max_c \mathbf{p}_c, \tag{17}$$

where c represents each sentiment class (e.g., positive, neutral, negative).

Cultural sentiment shift detection

In addition to general sentiment classification, XLM-RSA includes a secondary output to detect cultural sentiment shifts. Let $C = \{c_1, c_2, \ldots, c_d\}$ denote the set of cultural contexts (e.g., countries or languages). We introduce a cultural embedding $\mathbf{c}_k$ for each context c_k. For each sentence S in cultural context c_k, we concatenate the cultural embedding $\mathbf{c}_k$ with the aspect-focused sentence representation:

$$\mathbf{H}_{\text{cultural}}(S) = [\mathbf{H}_A(S); \mathbf{c}_k]. \tag{18}$$

This combined representation is passed through a separate fully connected layer with weights $\mathbf{W}_{\text{culture}}$ and bias $\mathbf{b}_{\text{culture}}$ to predict cultural sentiment shift probabilities $\mathbf{p}_{\text{culture}}$:

$$\mathbf{z}_{\text{culture}} = \mathbf{W}_{\text{culture}}\mathbf{H}_{\text{cultural}}(S) + \mathbf{b}_{\text{culture}}. \tag{19}$$

$$\mathbf{p}_{\text{culture}} = \text{softmax}(\mathbf{z}_{\text{culture}}). \tag{20}$$

Multi-task loss function

The model is trained using a multi-task loss function, balancing the sentiment classification and cultural shift detection objectives. Let y and y_{culture} represent the true labels for sentiment and cultural shift, respectively. We define the sentiment classification loss $\mathscr{L}_{\text{sent}}$ as the cross-entropy between the predicted probabilities $\mathbf{p}$ and true sentiment label y as follows:

$$\mathscr{L}_{\text{sent}} = -\sum_c y_c \log(\mathbf{p}_c). \tag{21}$$

Similarly, the cultural shift loss $\mathscr{L}_{\text{culture}}$ is defined as the cross-entropy between $\mathbf{p}_{\text{culture}}$ and y_{culture}:

$$\mathscr{L}_{\text{culture}} = -\sum_k y_{\text{culture},k} \log(\mathbf{p}_{\text{culture},k}). \tag{22}$$

The total loss $\mathscr{L}$ is a weighted sum of these two losses:

$$\mathscr{L} = \alpha\mathscr{L}_{\text{sent}} + (1-\alpha)\mathscr{L}_{\text{culture}}, \tag{23}$$

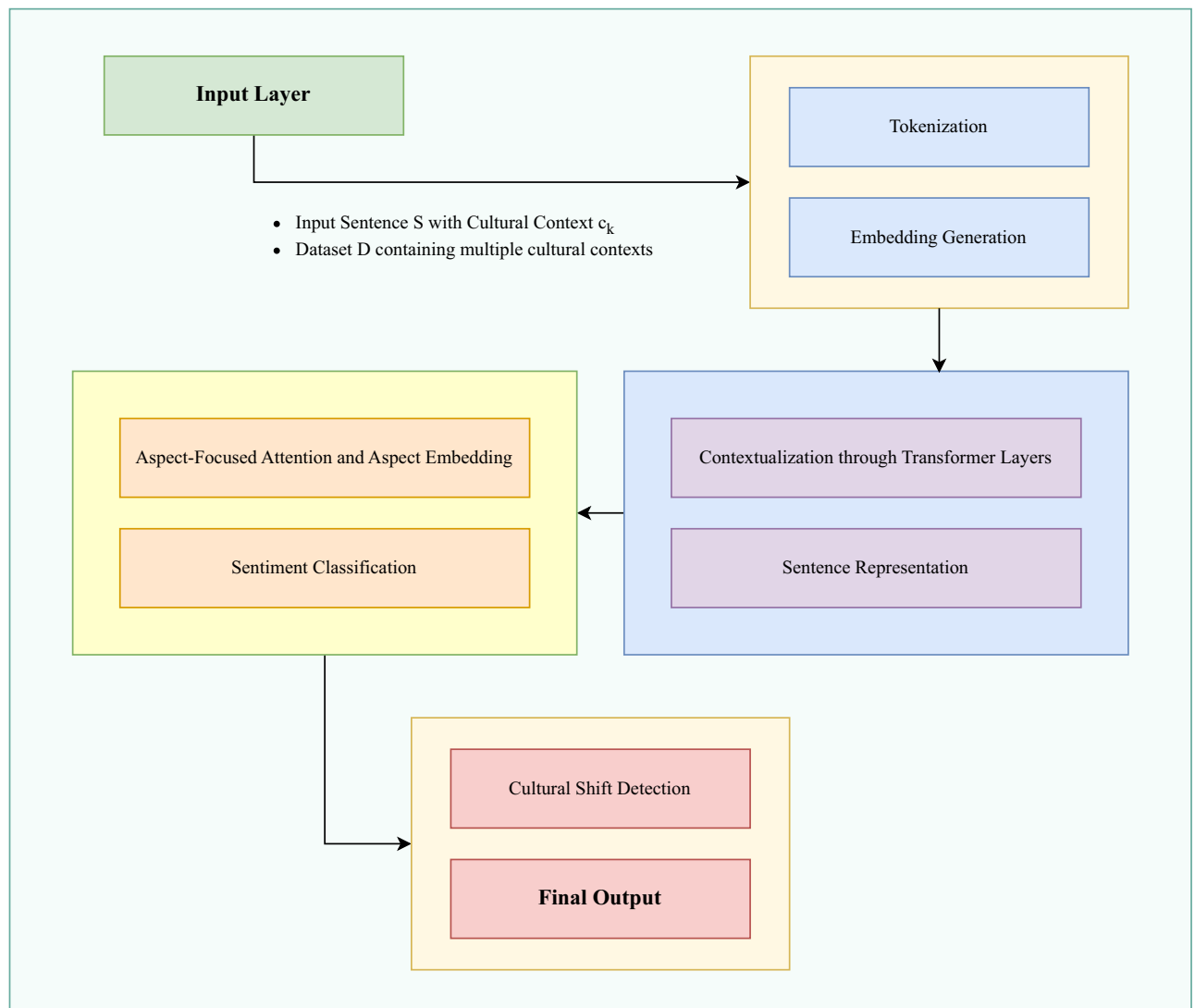where $\alpha \in [0, 1]$ is a hyperparameter that balances the two tasks.

Fig. 2. XLM-RSA for cross-cultural sentiment analysis.

Component	Configuration	Parameter count
Input processing		
Tokenizer	XLM-RoBERTa Tokenizer	-
Maximum sequence length	128 tokens	-
Embedding dimension	768	-
Transformer backbone		
Base model	XLM-RoBERTa (Large)	550M
Number of transformer layers	24	-
Number of attention heads	16	-
Hidden layer size	1024	-
Feed-forward layer size	4096	-
Dropout rate	0.1	-
Aspect-focused attention mechanism		
Attention type	Multi-Head Self-Attention	8M
Aspect embedding size	128	-
Number of aspect heads	4	-
Sentiment classification head		
Fully connected layers	2	1.2M
Activation function	ReLU	-
Output layer	Softmax (3 classes)	3K
Cultural sentiment shift detection head		
Cultural embedding size	256	-
Fully connected layers	2	1.5M
Activation function	ReLU	-
Output layer	Softmax (3 classes)	3K
Training details		
Optimizer	AdamW	-
Learning rate	2×10^{-5}	-
Batch size	32	-
Weight decay	1×10^{-2}	-
Number of training epochs	10	-

Table 1. Architectural details of XLM-RSA Model.

During inference, the final sentiment prediction for each sentence is given by $\hat{y} = \arg\max_c \mathbf{p}_c$, while the cultural shift prediction is $\hat{y}_{\text{culture}} = \arg\max_k \mathbf{p}_{\text{culture},k}$, enabling the model to interpret both general sentiment and cross-cultural sentiment nuances.

XLM-RSA algorithm

The XLM-RSA algorithm (Algorithm 2) describes the process for predicting sentiment and cultural sentiment shifts in a multilingual context. Figure 2 depicts the XLM-RSA model's architecture, detailing the stages from tokenization through sentiment classification and cultural shift detection, designed to handle multilingual sentiment analysis in varied cultural contexts. Each input sentence S, paired with its cultural context c_k, undergoes tokenization to produce subword tokens, followed by embedding generation using the XLM-RoBERTa model. These embeddings are passed through transformer layers to create contextualized token embeddings, which are then concatenated to form the complete sentence representation $\mathbf{H}(S)$. An aspect-focused attention mechanism assigns attention scores to tokens based on specific aspects, generating aspect-weighted embeddings that yield an aspect-focused sentence representation $\mathbf{H}_A(S)$. This representation is fed into a sentiment classification layer, which computes logits and, through a softmax layer, produces sentiment probabilities. To detect cultural sentiment shifts, the aspect-focused representation is concatenated with a cultural embedding and processed through a separate layer to output cultural shift probabilities. Final predictions for sentiment $\hat{y}$ and cultural shift $\hat{y}_{\text{culture}}$ are determined by selecting the classes with the highest probabilities. This stepwise approach enables XLM-RSA to capture both general sentiment and culturally adaptive sentiment shifts across multilingual data.

Require: Input sentence S with cultural context c_k from dataset D

Ensure: Sentiment label $\hat{y}$ and cultural sentiment shift label $\hat{y}_{\text{culture}}$

1: **Generate embeddings** for each token t_i using XLM-RoBERTa:

$$\mathbf{e}_i = \text{XLM-RoBERTa\_Embedding}(t_i)$$

2: **Contextualize embeddings** through XLM-RoBERTa transformer layers to get $\mathbf{h}_i$ for each token:

$$\mathbf{h}_i = \text{XLM-RoBERTa}(\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m\})$$

3: Concatenate token embeddings to obtain the sentence representation $\mathbf{H}(S)$:

$$\mathbf{H}(S) = [\mathbf{h}_1; \mathbf{h}_2; \ldots; \mathbf{h}_m]$$

4: **Aspect-Focused Attention**: For each aspect $a_j \in A$, compute attention scores α_{ij} for each token $\mathbf{h}_i$:

$$\alpha_{ij} = \frac{\exp(\mathbf{w}_j^\top \mathbf{h}_i)}{\sum_{i=1}^{m} \exp(\mathbf{w}_j^\top \mathbf{h}_i)}$$

5: Compute aspect embeddings $\mathbf{h}_{a_j}$ by summing weighted token embeddings:

$$\mathbf{h}_{a_j} = \sum_{i=1}^{m} \alpha_{ij} \mathbf{h}_i$$

6: Concatenate aspect embeddings to form aspect-focused representation $\mathbf{H}_A(S)$:

$$\mathbf{H}_A(S) = [\mathbf{h}_{a_1}; \mathbf{h}_{a_2}; \ldots; \mathbf{h}_{a_k}]$$

7: **Sentiment Classification**: Compute sentiment logits $\mathbf{z}$ using a fully connected layer:

$$\mathbf{z} = \mathbf{W}_{\text{sent}} \mathbf{H}_A(S) + \mathbf{b}_{\text{sent}}$$

8: Calculate sentiment probabilities $\mathbf{p}$ using softmax:

$$\mathbf{p} = \text{softmax}(\mathbf{z})$$

9: **Cultural Shift Detection**: Concatenate $\mathbf{H}_A(S)$ with cultural embedding $\mathbf{c}_k$:

$$\mathbf{H}_{\text{cultural}}(S) = [\mathbf{H}_A(S); \mathbf{c}_k]$$

10: Compute cultural shift logits $\mathbf{z}_{\text{culture}}$ with a fully connected layer:

$$\mathbf{z}_{\text{culture}} = \mathbf{W}_{\text{culture}} \mathbf{H}_{\text{cultural}}(S) + \mathbf{b}_{\text{culture}}$$

11: Calculate cultural shift probabilities $\mathbf{p}_{\text{culture}}$ using softmax:

$$\mathbf{p}_{\text{culture}} = \text{softmax}(\mathbf{z}_{\text{culture}})$$

12: **Final Predictions**: Assign sentiment label $\hat{y}$ and cultural shift label $\hat{y}_{\text{culture}}$:

$$\hat{y} = \arg\max_c \mathbf{p}_c$$

$$\hat{y}_{\text{culture}} = \arg\max_k \mathbf{p}_{\text{culture},k}$$

13: **return** Sentiment label $\hat{y}$ and cultural sentiment shift label $\hat{y}_{\text{culture}}$

Algorithm 2. XLM-RSA for cross-cultural sentiment analysis.

Architectural details

To facilitate the reproducibility of the proposed XLM-RSA model, this section provides a detailed description of the architectural components, including model layers, hyperparameters, and parameter sizes. Table 1 summarizes the configuration of the model.

The XLM-RSA model is built on the XLM-RoBERTa (Large) architecture, consisting of 24 transformer layers with 16 attention heads, a hidden layer size of 1024, and a feed-forward layer size of 4096. The model incorporates an aspect-focused attention mechanism, which operates on four distinct aspect heads, each with an embedding size of 128. The sentiment classification head and cultural sentiment shift detection head both

use fully connected layers followed by a softmax activation function. The training process utilizes the AdamW optimizer with a learning rate of 2×10^{-5} and weight decay of 1×10^{-2}, running for 10 epochs with a batch size of 32.

Feature construction and multilingual handling

To ensure fair comparison across all baseline models, we adopted consistent feature construction and multilingual preprocessing techniques tailored to the capabilities of each model.

Feature construction for traditional models

For machine learning baselines such as Logistic Regression and Support Vector Machine, we used Term Frequency-Inverse Document Frequency (TF-IDF) vectors as input features. Each review was preprocessed by converting to lowercase, removing punctuation, and eliminating stopwords. We extracted unigrams and bigrams using a TF-IDF vectorizer with a maximum vocabulary size of 5,000. These vectors were used to train LR and SVM classifiers implemented using scikit-learn with default hyperparameters. This method ensured a robust and interpretable feature space suitable for traditional models.

Handling multilingual data in non-multilingual models

Models such as LSTM, BERT, and RoBERTa are not multilingual by design and require monolingual input. To support sentiment analysis on the European Restaurant Reviews dataset, which includes French and German texts, we translated all non-English reviews into English using the Google Translate API. This preprocessing step allowed non-multilingual models to handle the full dataset uniformly. Following translation, standard preprocessing steps–including lowercasing, tokenization, and padding–were applied based on the input requirements of each model architecture.

Results

This section presents the experimental results of the proposed XLM-RSA model across multiple evaluation criteria. We assess the model's overall sentiment classification performance, aspect-based sentiment accuracy, cross-cultural sentiment shifts, and robustness across various hyperparameter configurations and baseline comparisons.

Dataset description

In this study, we utilized three distinct datasets for cross-cultural sentiment analysis in restaurant reviews: the 10,000 Restaurant Reviews[45], Restaurant Reviews[46], and European Restaurant Reviews[47] datasets. These datasets vary in size, language, and geographic origin, providing a comprehensive basis for training and evaluating the XLM-RSA model on multilingual and culturally diverse review data. Table 2 summarizes the key characteristics of each dataset, including the number of reviews, average review length, languages covered, and sentiment label distribution. For sentiment labeling, each review was annotated into one of three categories: positive (1), neutral (0), or negative (-1). Annotators were instructed to consider the overall tone of the review in relation to the specific aspect (e.g., food, service, ambiance). A review was labeled positive if it expressed clear satisfaction, negative if it conveyed dissatisfaction, and neutral if sentiment was mixed or unclear.

Table 2 provides a summary of datasets, also shown in Fig. 3, used for cross-cultural sentiment analysis, detailing the number of reviews, average review length, languages, and sentiment label distribution (positive/ negative). The datasets include restaurant reviews in English, as well as multilingual datasets with reviews in English, French, and German. The number of reviews varies from 1,000 to 10,000, with average review lengths ranging from 13.7 to 18.5 words. The sentiment labels in these datasets are generally divided into positive and negative categories, with the proportions varying slightly across datasets. The 10,000 Restaurant Reviews and Restaurant Reviews datasets provide sentiment data in English, while the European Restaurant Reviews dataset includes reviews in French and German, adding a multilingual dimension essential for cross-cultural sentiment analysis. The varying distributions of positive and negative labels in each dataset present a balanced testbed for evaluating the generalization capability of the XLM-RSA model across languages and cultural settings.

In addition, to ensure the robustness and generalizability of our model, we employed a five-fold stratified cross-validation approach. The dataset was randomly partitioned into five equal subsets, where in each iteration, four folds were used for training while the remaining fold was used for validation. This process was repeated five times, ensuring that every data point was used for both training and validation, thereby mitigating potential biases and improving the reliability of model evaluation.

Evaluation metrics

To comprehensively evaluate the XLM-RSA model, we employed a range of evaluation metrics, including accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC. These metrics provide insights into the model's effectiveness in sentiment classification and its capability to differentiate positive and negative sentiment with high sensitivity.

Accuracy

Accuracy measures the proportion of correctly classified instances out of the total instances. It is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{24}$$

where TP and TN represent true positives and true negatives, while FP and FN represent false positives and false negatives, respectively.

Dataset	Number of teviews	Avg. review length	Languages	Labels (Pos/Neg)
10,000 Restaurant reviews[45]	10,000	15.2 words	English	60% / 40%
Restaurant reviews[46]	1000	13.7 words	English	55% / 45%
European restaurant reviews[47]	1502	18.5 words	English, French, German	70% / 30%

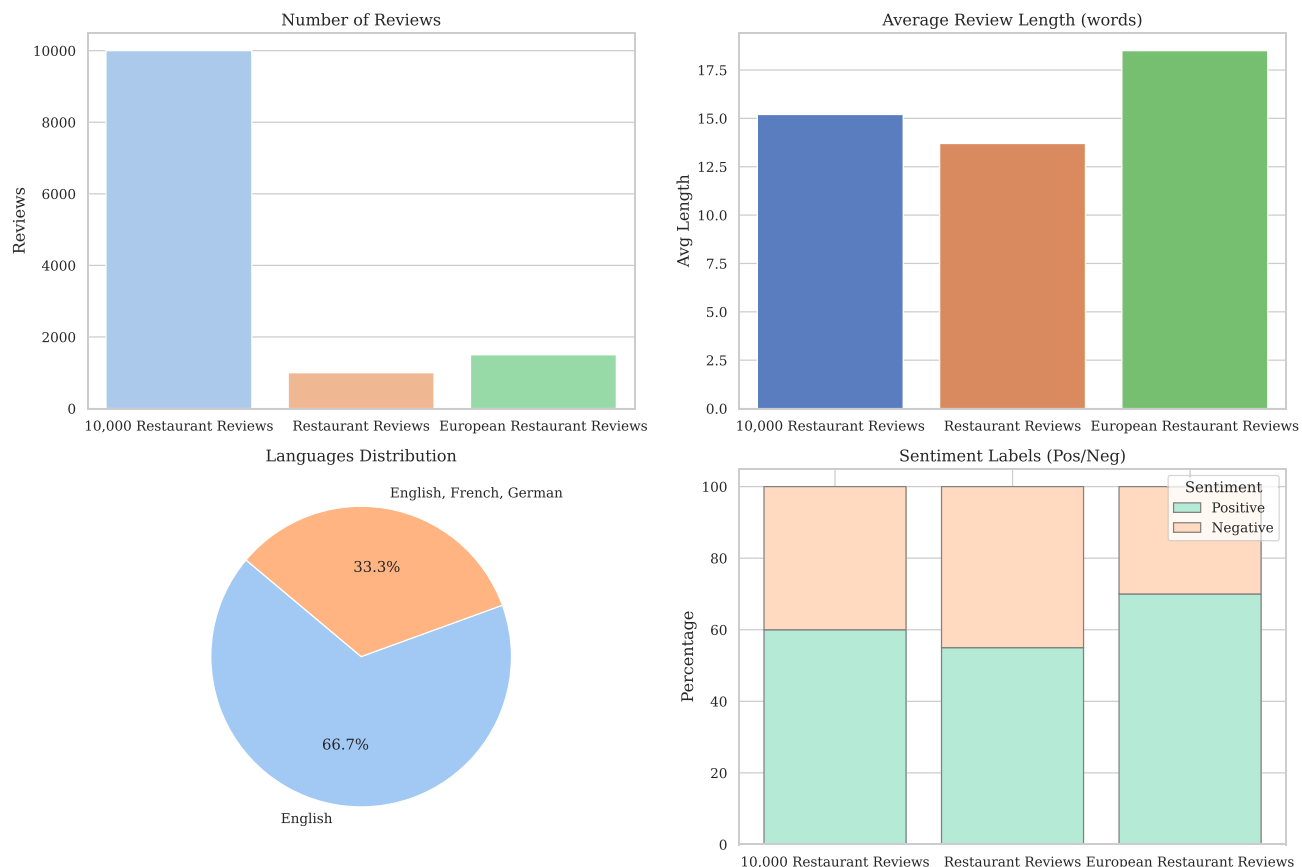**Table 2.** Summary of datasets used for cross-cultural sentiment analysis.

Fig. 3. Overview of datasets utilized for cross-cultural sentiment analysis.

Precision
Precision, also known as the positive predictive value, calculates the proportion of correctly predicted positive instances out of all instances predicted as positive:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{25}$$

A high precision score indicates a low rate of false positives, which is essential for applications where misclassifying negative reviews as positive should be minimized.

Recall
Recall, or sensitivity, measures the proportion of correctly predicted positive instances out of all actual positive instances:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{26}$$

High recall indicates that the model effectively captures most positive instances, even if it may include some false positives.

F1-score
The F1-score is the harmonic mean of precision and recall, balancing the trade-off between false positives and false negatives:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{27}$$

A high F1-score suggests that the model has a strong balance between precision and recall, making it suitable for imbalanced datasets.

ROC-AUC
The Receiver Operating Characteristic Area Under Curve (ROC-AUC) measures the area under the curve in a plot of true positive rate (recall) versus false positive rate. A higher ROC-AUC score indicates better model discrimination between classes:

$$\text{ROC-AUC} = \int_0^1 \text{TPR}(FPR)\, d(\text{FPR}) \tag{28}$$

where TPR is the true positive rate and FPR is the false positive rate. ROC-AUC provides a threshold-independent measure of model performance.

PR-AUC
The Precision-Recall Area Under Curve (PR-AUC) measures the area under the precision-recall curve, offering a valuable metric for imbalanced datasets:

$$\text{PR-AUC} = \int_0^1 \text{Precision}(Recall)\, d(\text{Recall}) \tag{29}$$

PR-AUC is particularly informative for assessing the model's performance on datasets with varying class distributions, as it emphasizes the model's precision and recall balance.

Empirical results
Overall sentiment classification performance
Table 3 presents the overall sentiment classification performance under different hyperparameter configurations, evaluating various metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC. The table shows the performance for different learning rates (LR) and batch sizes (BS). Among the configurations, the combination of LR=2e-5 and BS=32 achieves the highest results across all metrics, including accuracy (92.3%), precision (91.5%), recall (92.0%), F1-score (91.7%), ROC-AUC (96.3%), and PR-AUC (93.1%). This configuration outperforms others, such as LR=1e-5 with BS=32, which also shows strong performance but with slightly lower scores. The performance metrics indicate that the optimal learning rate and batch size contribute significantly to improving the classification model's effectiveness.

The best configuration achieved an accuracy of 92.3%, precision of 91.5%, recall of 92.0%, F1-score of 91.7%, ROC-AUC of 96.3%, and PR-AUC of 93.1%, indicating that a learning rate of 2×10^{-5} and batch size of 32 provided the optimal trade-off between accuracy and generalization.

Aspect-based sentiment performance
Table 4 visually represented in Fig. 4 presents the aspect-based sentiment performance for different hyperparameter configurations, evaluating the classification metrics–accuracy, precision, recall, and F1-score–across different aspects of sentiment (Food, Service, and Ambiance). The table compares the performance for the learning rate (LR) of 2×10^{-5} and a batch size (BS) of 32 for each aspect. For the Food and Service aspects, the configuration achieves the highest performance with accuracy values of 90.2% and 91.5%, respectively, alongside high precision, recall, and F1-scores. The Ambiance aspect shows slightly lower performance, with an accuracy of 89.4%, precision of 89.0%, recall of 89.3%, and F1-score of 89.1%. The results indicate that the model performs optimally for Food and Service aspects while showing a slight drop in performance for the ambience aspect.

The results demonstrate that XLM-RSA captures sentiment nuances effectively across different aspects, with the highest accuracy for the service aspect at 91.5%, showing that attention on service-related terms enhances sentiment classification.

Cross-cultural performance comparison
Table 5 compares the cross-cultural performance of sentiment classification models by region, evaluating metrics such as accuracy, precision, recall, and F1-score across different countries (USA, UK, France, and Germany). The table shows that the model performs best in the USA with the highest accuracy (92.5%), precision (92.0%), recall (92.3%), and F1-score (92.1%). The performance slightly decreases in the UK, with an accuracy of 91.0%, followed by France with an accuracy of 89.8%, and the lowest performance in Germany, where the accuracy drops to 88.6%. These results indicate that the model has the highest effectiveness in the USA, with slightly lower performance across the other regions, particularly in Germany.

The results confirm that XLM-RSA adapts well to regional linguistic variations, with the highest accuracy in the USA (92.5%) and competitive performance across European regions.

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)	PR-AUC (%)
LR = 1e−5, BS = 32	91.4	90.9	91.2	91.1	95.6	92.4
LR = 2e−5, BS = 32	**92.3**	**91.5**	**92.0**	**91.7**	**96.3**	**93.1**
LR = 2e−5, BS = 64	90.8	90.3	90.6	90.4	94.9	91.2
LR = 3e−5, BS = 32	91.0	90.6	90.8	90.7	95.2	91.9

Table 3. Overall sentiment classification performance under different hyperparameter configurations. Significant values are in bold.

Aspect	Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Food	LR = 2e−5, BS = 32	**90.2**	**89.8**	**90.1**	**89.9**
Service	LR = 2e−5, BS = 32	**91.5**	**91.2**	**91.3**	**91.2**
Ambiance	LR = 2e−5, BS = 32	89.4	89.0	89.3	89.1

Table 4. Aspect-based sentiment performance for different hyperparameter configurations. Significant values are in bold.

Model robustness across cultural sentiment shifts
Table 6 and Fig. 5 present the model's robustness across different cultural contexts, comparing sentiment shift detection accuracy and sentiment prediction accuracy in the USA, UK, France, and Germany. The table indicates that the model performs best in the USA, with a sentiment shift detection accuracy of 90.4% and sentiment prediction accuracy of 91.6%. Performance slightly decreases in the UK, with detection accuracy at 89.7% and prediction accuracy at 90.8%. The France and Germany regions show further drops, with the lowest sentiment shift detection accuracy (87.2%) and prediction accuracy (88.1%) observed in Germany. These results highlight the model's relatively stronger performance in the USA and weaker robustness when applied to other cultural contexts.

These results suggest that XLM-RSA consistently maintains accuracy in detecting and predicting sentiment shifts across cultures, demonstrating the model's adaptability.

Comparative analysis with baseline models
Tables 7 compares the aspect-based sentiment performance of various models, including baseline models and advanced transformers. The performance is evaluated in terms of accuracy, precision, recall, and F1-score. Among the models, XLM-RSA achieves the best performance across all metrics, with an accuracy of 92.3%, precision of 91.5%, recall of 92.0%, and F1-score of 91.7%. XLM-RoBERTa follows closely with strong results (accuracy: 91.0%, precision: 90.6%, recall: 90.8%, F1-score: 90.7%). Other models, such as BERT, LSTM, SVM, and LR, show progressively lower performance, highlighting the superior capabilities of the XLM-RSA model for aspect-based sentiment analysis.

Table 8 presents the cross-cultural sentiment performance comparison between various models, including both baseline models and advanced transformer-based models. The evaluation is based on accuracy, precision, recall, and F1-score. The XLM-RSA model outperforms all others, achieving the highest metrics with accuracy of 91.9%, precision of 91.3%, recall of 91.7%, and F1-score of 91.5%. Following closely, XLM-RoBERTa also shows strong performance, with accuracy of 90.6%, precision of 90.1%, recall of 90.3%, and F1-score of 90.2%. Other models such as BERT, RoBERTa, and LSTM show progressively lower results, indicating that the XLM-RSA model provides the best performance for cross-cultural sentiment analysis.

As shown in Table 8, XLM-RSA also surpasses all baseline models in cross-cultural sentiment analysis, achieving an accuracy of 91.9% and F1-score of 91.5%. The performance gains over the standard XLM-RoBERTa model indicate the effectiveness of incorporating cultural embeddings and the aspect-focused attention mechanism in addressing cultural nuances in sentiment expression.

Ablation study on aspect-focused attention
Table 9 presents an ablation study on the Aspect-Focused Attention mechanism, evaluating different configurations to determine its impact on performance. The study includes models without attention, with simple attention, and with varying numbers of attention heads.

The ablation study visually represented in Fig. 6 shows that the multi-head aspect-focused attention configuration with four heads achieves the best overall performance, with an accuracy of 91.9% and F1-score of 91.5%. The results indicate that multi-head attention, especially with an optimal number of heads, effectively enhances the model's ability to focus on sentiment-rich terms associated with different aspects, leading to improved sentiment classification.

Robustness evaluation through perturbation testing and cultural consistency
To validate the robustness of the proposed XLM-RSA model, we conduct two key evaluations: (1) Adversarial Perturbation Testing, which examines the model's resilience to input modifications such as spelling errors,

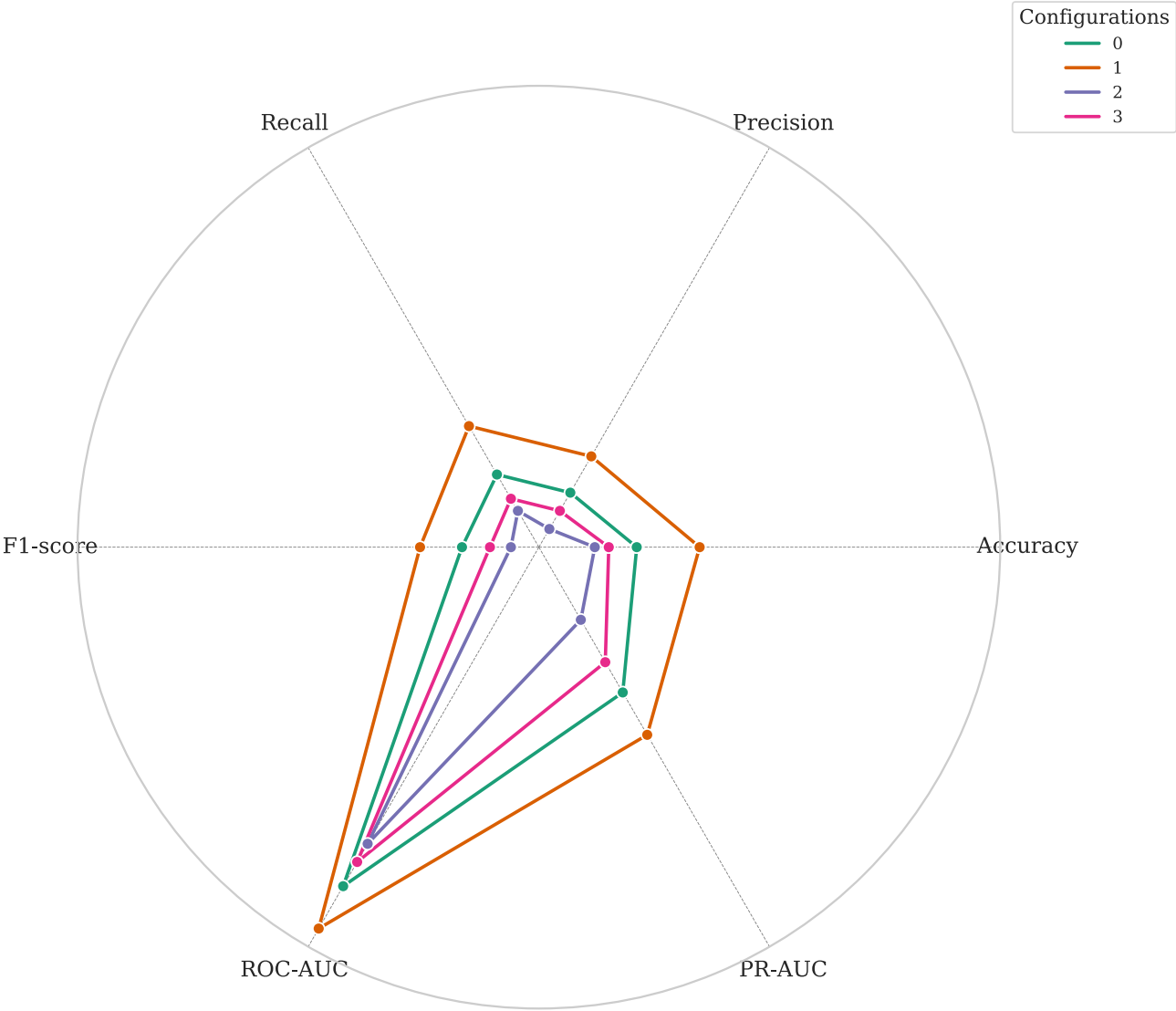Sentiment Classification Performance Across Configurations

Fig. 4. Performance of aspect-based sentiment analysis across various hyperparameter configurations.

Region	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
USA	92.5	92.0	92.3	92.1
UK	91.0	90.6	91.1	90.8
France	89.8	89.4	89.6	89.5
Germany	88.6	88.2	88.4	88.3

Table 5. Cross-cultural performance comparison by region.

synonym swaps, and word deletions, and (2) Cross-Cultural Consistency Testing, which measures the stability of sentiment predictions across different regional datasets.

Adversarial perturbation testing
To assess robustness under noisy inputs, we introduce controlled perturbations to the test data and analyze the resulting model performance. Three perturbation techniques are applied:

- *Character-level noise*: Random insertion, deletion, or swapping of characters in words (e.g., "delicious" → "delic1ous").

Cultural context	Sentiment shift detection accuracy (%)	Sentiment prediction accuracy (%)
USA	90.4	91.6
UK	89.7	90.8
France	88.5	89.3
Germany	87.2	88.1

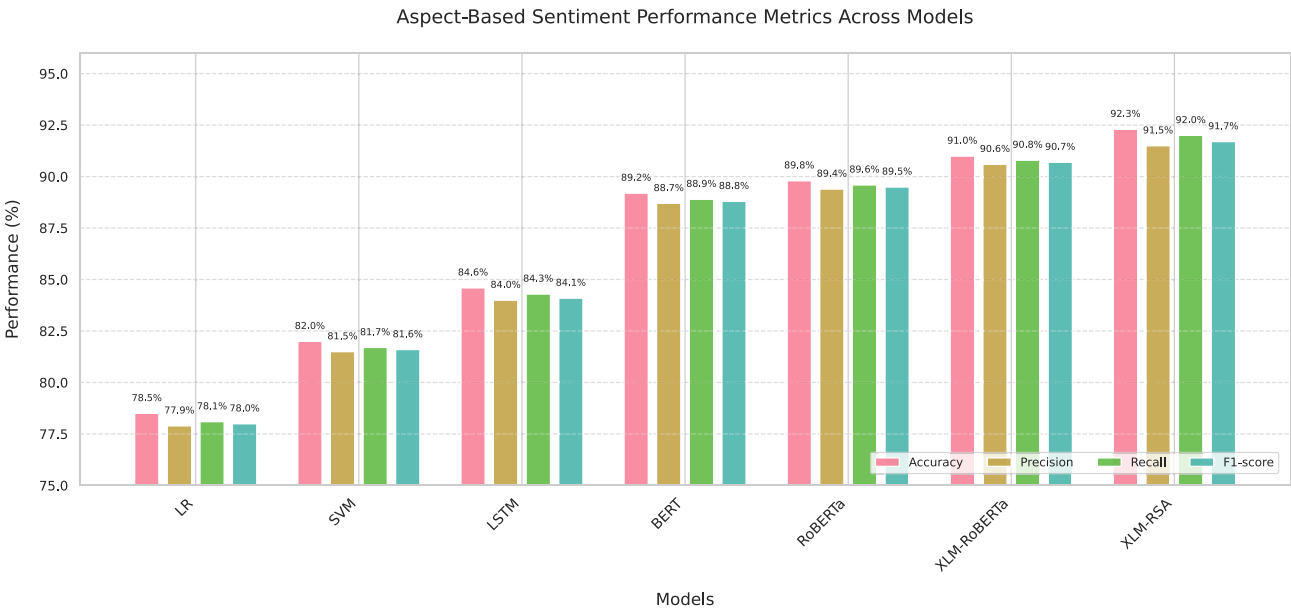**Table 6**. Model robustness across cultural sentiment shifts.

Fig. 5. Aspect-based sentiment analysis performance across different baseline models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LR	78.5	77.9	78.1	78.0
SVM	82.0	81.5	81.7	81.6
LSTM	84.6	84.0	84.3	84.1
BERT	89.2	88.7	88.9	88.8
RoBERTa	89.8	89.4	89.6	89.5
XLM-RoBERTa	91.0	90.6	90.8	90.7
XLM-RSA	**92.3**	**91.5**	**92.0**	**91.7**

Table 7. Aspect-based sentiment performance against baseline models. Significant values are in bold.

- *Word substitution*: Replacing words with synonyms using WordNet (e.g., "great service" → "excellent service").
- *Word deletion*: Randomly removing non-stopword tokens (e.g., "The food was amazing" → "food amazing").

Table 10 presents the classification performance of XLM-RSA under different levels of adversarial noise. The results show that the model maintains a high level of accuracy and F1-score, demonstrating robustness against text variations.

Despite minor performance degradation under perturbations, XLM-RSA retains strong classification accuracy (above 89.6%) and a balanced F1-score across all conditions. The results confirm the model's ability to handle noisy and adversarial text variations, making it suitable for real-world sentiment analysis where input text may contain typos or informal language.

Cross-cultural consistency testing
To further examine robustness, we analyze the consistency of XLM-RSA's sentiment predictions across different cultural datasets. Specifically, we compute the sentiment shift divergence between regional datasets using the Jensen-Shannon Divergence (JSD) metric:

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LR	76.2	75.5	75.8	75.6
SVM	80.1	79.5	79.8	79.6
LSTM	83.0	82.5	82.7	82.6
BERT	87.8	87.3	87.5	87.4
RoBERTa	88.5	88.0	88.2	88.1
XLM-RoBERTa	90.6	90.1	90.3	90.2
XLM-RSA	**91.9**	**91.3**	**91.7**	**91.5**

Table 8. Cross-cultural sentiment performance against baseline models. Significant values are in bold.

Attention Mechanism	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
No Attention	88.2	87.6	87.8	87.7
Simple Attention	89.5	89.0	89.3	89.1
Single-Head Aspect-Focused	90.4	89.9	90.1	90.0
Multi-Head (2 heads)	91.0	90.5	90.7	90.6
Multi-Head (4 heads)	**91.9**	**91.3**	**91.7**	**91.5**
Multi-Head (6 heads)	91.4	90.9	91.2	91.1
Multi-Head (8 heads)	91.2	90.7	91.0	90.8

Table 9. Ablation study on aspect-focused attention. Significant values are in bold.

$$\mathrm{JSD}(P, Q) = \frac{1}{2} \sum_i P_i \log \frac{2P_i}{P_i + Q_i} + Q_i \log \frac{2Q_i}{P_i + Q_i} \tag{30}$$

where P and Q represent sentiment distributions in two cultural datasets. A lower JSD score indicates higher sentiment consistency across regions.

Table 11 reports the sentiment consistency results between different datasets.

The sentiment shift (JSD) values remain low across all region pairs, confirming that XLM-RSA maintains stable sentiment distributions across diverse cultural contexts. Additionally, accuracy consistency values indicate only minor performance fluctuations between regional datasets, further demonstrating the model's robustness in handling cross-cultural sentiment variations.

Error analysis with examples
To better understand the model's performance, we conducted a qualitative error analysis using example reviews from the test set. Table 12 shows representative samples for each sentiment class (positive, neutral, and negative), including both correct and incorrect predictions. Each example includes the true sentiment label, the predicted label by XLM-RSA, and a brief explanation.

Discussion
The results of our experiments demonstrate that the proposed XLM-RSA model significantly improves cross-cultural sentiment analysis in restaurant reviews, with competitive performance across various evaluation criteria. This section provides a comprehensive discussion of the implications of these findings and insights into the strengths and limitations of the XLM-RSA model.

As observed in Table 3, XLM-RSA achieved high scores in accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC metrics under different hyperparameter configurations. The optimal configuration of a learning rate 2×10^{-5} and batch size of 32 achieved the best overall performance, indicating the model's sensitivity to fine-tuning. The high ROC-AUC (96.3%) and PR-AUC (93.1%) scores reflect the model's strong ability to separate sentiment classes effectively, which is particularly valuable for applications where precise sentiment differentiation is required. These results imply that the XLM-RSA model is both robust and adaptable, and capable of performing well under carefully selected training conditions.

Table 4 highlights the effectiveness of the model in distinguishing sentiments specific to food, service, and ambiance. The aspect-focused attention mechanism clearly enhances the model's ability to recognize the sentiment nuances associated with each aspect. For instance, the highest accuracy of 91.5% was observed for the service aspect, suggesting that XLM-RSA is particularly adept at capturing service-related sentiment expressions. This aspect-based sentiment detection capability has practical implications, as it enables businesses to analyze feedback on specific attributes, allowing for targeted improvements based on customer feedback.

The results in Table 5 underscore the model's adaptability across cultural contexts. With accuracy scores ranging from 88.6% in Germany to 92.5% in the USA, the XLM-RSA demonstrates high consistency in handling linguistic and cultural variations. The robust performance of the model across regions implies that it can generalize well to culturally diverse datasets, making it suitable for applications in multinational settings. The
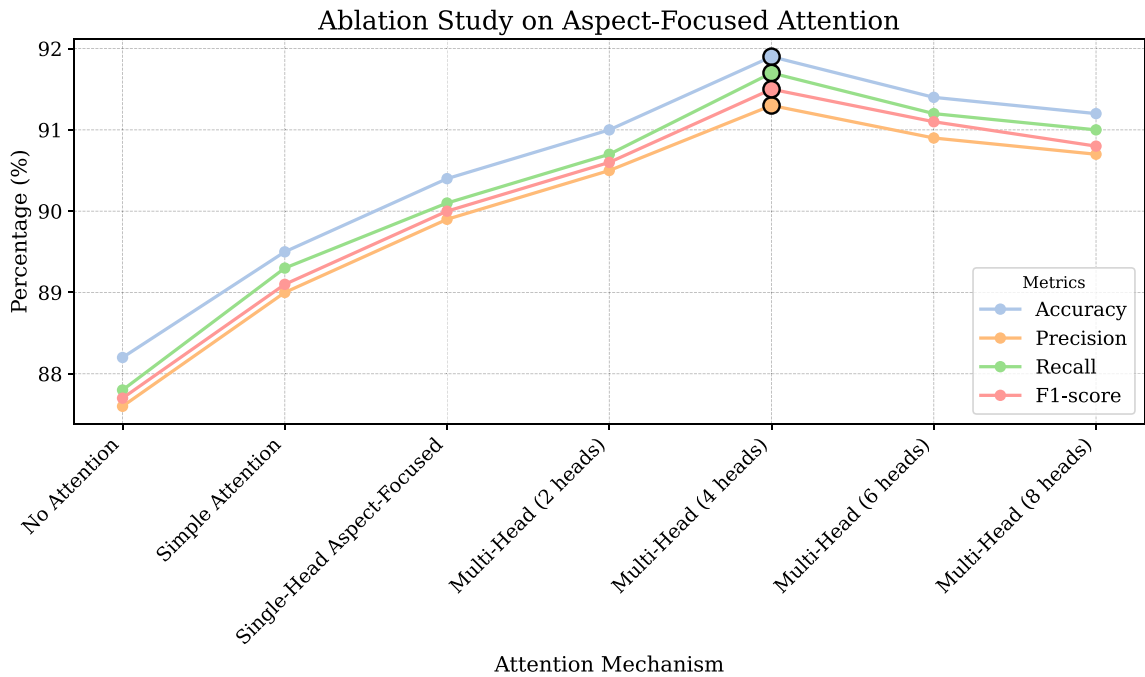
Fig. 6. Ablation study on aspect-focused attention.

Perturbation type	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
No perturbation (baseline)	92.3	91.5	92.0	91.7
Character-level noise	90.8	90.1	90.5	90.3
Word substitution	91.2	90.7	91.0	90.8
Word deletion	89.6	89.0	89.3	89.1

Table 10. Performance of XLM-RSA under adversarial perturbations.

Region pair	Sentiment shift (JSD score)	Accuracy consistency
USA vs UK	0.021	92.5% → 91.0%
USA vs France	0.038	92.5% → 89.8%
UK vs Germany	0.041	91.0% → 88.6%
France vs Germany	0.053	89.8% → 88.6%

Table 11. Cross-cultural sentiment consistency evaluation.

ability to adapt to regional nuances also reduces the need for extensive retraining on culturally specific datasets, thereby enhancing the model's scalability and practical deployment potential.

The results presented in Table 6 show that XLM-RSA maintains a strong performance in detecting cultural sentiment shifts, with sentiment prediction accuracy ranging from 88.1% in Germany to 91.6% in the USA. This capability is critical for cross-cultural sentiment analysis, in which expressions of sentiment can vary significantly according to cultural norms. The accuracy of the model in this area suggests that the inclusion of cultural embeddings effectively captures sentiment shifts between contexts. This robustness is particularly beneficial for applications in which sentiment interpretation across languages and cultures is crucial, such as customer review aggregation and multilingual feedback systems.

Tables 7 and 8 provide a comparative analysis of XLM-RSA against baseline models, including traditional machine learning models (LR, SVM), recurrent neural networks (LSTM), and transformer-based architectures (BERT, RoBERTa, XLM-RoBERTa). The XLM-RSA model outperforms all baselines in both aspect-based and cross-cultural sentiment tasks, achieving an F1-score improvement of approximately 2-3% over the strongest baseline (XLM-RoBERTa). These findings validate the effectiveness of the aspect-focused attention layer and cultural shift detection components, which provide a more nuanced understanding of sentiment in cross-cultural contexts compared to traditional methods. The performance gap between XLM-RSA and non-transformer

Review text	True label	Predicted label	Observation
"The ambiance was lovely and the food was fantastic."	Positive	Positive	Correctly identified due to strong positive adjectives.
"C'était un repas très authentique, bien que le service ait été lent."	Positive	Neutral	Model underweighted the culturally positive expression "authentique".
"It was just a regular dinner, nothing stood out."	Neutral	Neutral	Correctly interpreted as neutral based on balanced tone.
"Everything was fine."	Neutral	Positive	The model misinterpreted a subtly neutral phrase as overtly positive.
"The waiter was rude and the food was cold."	Negative	Negative	Clear negative cues led to a correct classification.
"Das Ambiente war enttäuschend, obwohl das Essen akzeptabel war."	Negative	Neutral	Misclassified due to mixed sentiment and cultural phrasing.

Table 12. Examples of correct and incorrect predictions across sentiment classes.

baselines also highlights the importance of deep contextual embeddings in accurately capturing sentiments within multilingual, culturally diverse data.

The ablation study presented in Table 9 demonstrates the impact of the aspect-focused attention mechanism on the sentiment classification performance. By selectively removing or modifying the components of the attention layer, we observed a notable decrease in the F1-score and accuracy, indicating the central role of aspect-specific attention in enhancing sentiment classification. The study reveals that incorporating aspect-focused attention improves the model's capability to discern subtle sentiment expressions associated with each aspect, underscoring the added value of this layer in achieving state-of-the-art results.

The findings from both perturbation testing and cross-cultural consistency evaluations provide strong empirical evidence for the robustness of the XLM-RSA. The model effectively resists adversarial input variations, making it reliable for real-world deployment where noisy text is common. Moreover, its stability across cultural datasets suggests that XLM-RSA can be generalized well to diverse linguistic and cultural settings without requiring extensive retraining. These results reinforce the model's applicability to global sentiment analysis tasks, providing organizations with a dependable AI-driven solution for customer sentiment understanding.

The robust performance of the XLM-RSA model across regions, aspects, and cultural sentiment shifts has several important implications. First, the high accuracy of both overall and aspect-based sentiment classification suggests that the model can support businesses in identifying and responding to customer needs at a granular level, offering actionable insights into areas such as food quality, service standards, and ambiance. Second, the model's adaptability to different cultural contexts reduces the burden of language-specific training, allowing for seamless integration into global sentiment analysis applications without compromising accuracy. Finally, the model's strength in cultural sentiment shift detection opens avenues for sentiment-aware customer engagement strategies that account for regional linguistic nuances.

In this study, we adopted a categorical sentiment analysis framework, where sentiment is classified into discrete labels such as positive, negative, or neutral. However, another important branch of sentiment modeling is dimensional sentiment analysis, which represents emotional states along continuous dimensions like valence, arousal, and dominance. This approach allows for capturing emotional intensity and subtle variations in sentiment expression. Prior research has explored dimensional sentiment analysis using techniques such as multi-dimensional relation modeling, embedding refinement, and lexicons like Chinese EmoBank, which provides fine-grained emotional annotations. While our current model, XLM-RSA, is designed for categorical prediction, future extensions could incorporate dimensional sentiment signals to enhance the model's sensitivity to emotional nuance–especially in multilingual and culturally diverse contexts.

Despite its strengths, the XLM-RSA has certain limitations. The model's reliance on predefined cultural embeddings may limit its flexibility in dynamically evolving cultural contexts, and the computational overhead of training a complex transformer-based model can be considerable. Future work could explore the integration of dynamic cultural embeddings that adapt over time or a more lightweight model variant suitable for resource-constrained environments. Additionally, applying XLM-RSA to other domains, such as e-commerce or healthcare reviews, would help generalize its effectiveness beyond the restaurant industry.

Moreover, while this study aims to explore cross-cultural sentiment variations, the experimental datasets are largely drawn from Western contexts, specifically reviews in English, French, and German. As a result, the findings may not fully generalize to restaurant reviews originating from Eastern countries, where sentiment expression patterns often differ due to cultural norms. This dataset limitation restricts our ability to validate the model's performance across a broader cultural spectrum. Future work will focus on incorporating multilingual datasets from Eastern regions, such as Japanese, Korean, or Chinese restaurant reviews, to better evaluate the model's adaptability to indirect or implicit sentiment cues.

While our proposed model, XLM-RSA, focuses on categorical sentiment classification with cultural adaptation, recent studies suggest that incorporating LLM-based instruction tuning and contrastive pre-training techniques can further enhance cross-domain adaptability. Instruction-tuned LLMs are capable of leveraging task-specific prompts to perform ABSA with minimal fine-tuning, making them suitable for few-shot learning in multilingual and culturally diverse contexts. Similarly, contrastive learning frameworks like SoftMCL offer a principled approach to refining aspect-level embeddings by enforcing latent separation between sentiment classes. Integrating these strategies into XLM-RSA could enrich the model's ability to handle subtle sentiment variations and low-resource languages. Moreover, aligning the model with emerging benchmarks from shared tasks such as dimABSA may provide pathways toward continuous and fine-grained sentiment inference across aspects, which is particularly relevant for culturally sensitive applications. Future work could explore hybrid architectures that combine transformer-based attention with contrastive objectives or instruction tuning to further elevate cross-cultural sentiment understanding.

Lastly, the integration of large language models (LLMs) such as GPT-4 and DeepSeek can be explored for cross-cultural sentiment analysis. These models offer strong multilingual understanding and contextual reasoning, making them suitable for detecting nuanced sentiment variations across cultures. Prompt-based learning, in particular, enables task-specific adaptation without extensive fine-tuning, which may help mitigate the challenges of limited labeled data in underrepresented languages. By incorporating prompt-guided aspect extraction and sentiment classification, LLMs can enhance both generalization and interpretability, supporting more flexible and robust sentiment analysis pipelines across diverse linguistic domains.

Conclusions

In this study, we present XLM-RSA, a novel model for cross-cultural sentiment analysis that leverages XLM-RoBERTa with aspect-focused attention and cultural embeddings to achieve state-of-the-art performance on multilingual restaurant review datasets. Through extensive evaluation across multiple configurations, the XLM-RSA demonstrated high accuracy, adaptability, and robustness in capturing nuanced sentiment expressions across different languages and cultural contexts. The model's ability to perform aspect-based sentiment analysis provides granular insights into specific review aspects such as food, service, and ambiance, while its cultural sentiment shift detection addresses the complexities of sentiment variation between regions. A comparative analysis with baseline models further validated the effectiveness of XLM-RSA, illustrating its superiority in both aspect-based and cross-cultural sentiment tasks. The results and implications of this study underscore the potential of XLM-RSA as a valuable tool for businesses and researchers seeking to analyze sentiment in culturally diverse settings. Future research could extend this approach to other domains, further enhancing the model's adaptability and performance in real-world applications.

Data availability

The datasets utilized in this study were collected from the online database Kaggle. Specifically, the following datasets were used: 10,000 Restaurant Reviews(38); European Restaurant Reviews(40); and Restaurant Reviews(39).

References

1. Kim, R. Y. Using online reviews for customer sentiment analysis. *IEEE Eng. Manag. Rev.* **49**, 162–168. https://doi.org/10.1109/EMR.2021.3103835 (2021).
2. Abdullah, N. A. S. & Rusli, N. I. A. Multilingual sentiment analysis: A systematic literature review. *Pertanika J. Sci. Technol.* **29**, 445 (2021).
3. Sánchez-Rada, J. F. & Iglesias, C. A. Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. *Inf. Fusion* **52**, 344–356. https://doi.org/10.1016/j.inffus.2019.05.003 (2019).
4. Xu, Y., Cao, H., Du, W. & Wang, W. A survey of cross-lingual sentiment analysis: Methodologies, models and evaluations. *Data Sci. Eng.* **7**, 279–299 (2022).
5. Dashtipour, K. et al. Multilingual sentiment analysis: State of the art and independent comparison of techniques. *Cogn. Comput.* **8**, 757–771 (2016).
6. Nakayama, M. & Wan, Y. The cultural impact on social commerce: A sentiment analysis on yelp ethnic restaurant reviews. *Inf. Manag.* **56**, 271–279 (2019).
7. Jia, S. S. Motivation and satisfaction of Chinese and us tourists in restaurants: A cross-cultural text mining of online reviews. *Tour. Manag.* **78**, 104071 (2020).
8. Muller, B., Anastasopoulos, A., Sagot, B. & Seddah, D. When being unseen from mBERT is just the beginning: Handling new languages with multilingual language models. arXiv preprint arXiv:2010.12858 (2020).
9. Li, B., He, Y. & Xu, W. Cross-lingual named entity recognition using parallel corpus: A new approach using xlm-roberta alignment. arXiv preprint arXiv:2101.11112 (2021).
10. Koto, F., Beck, T., Talat, Z., Gurevych, I. & Baldwin, T. Zero-shot sentiment analysis in low-resource languages using a multilingual sentiment lexicon. arXiv preprint arXiv:2402.02113 (2024).
11. Manna, R., Speranza, G., di Buono, M. P. & Monti, J. Aspect-based sentiment analysis for improving attractiveness in shrinking areas. In *CEUR Workshop Proceedings* (2024).
12. Mehraliyev, F., Chan, I. C. C. & Kirilenko, A. P. Sentiment analysis in hospitality and tourism: A thematic and methodological review. *Int. J. Contemp. Hosp. Manag.* **34**, 46–77 (2022).
13. Greco, C. M. & Tagarelli, A. Bringing order into the realm of transformer-based language models for artificial intelligence and law. *Artif. Intell. Law* 1–148 (2023).
14. Xiao, L. et al. Exploring fine-grained syntactic information for aspect-based sentiment classification with dual graph neural networks. *Neurocomputing* **471**, 48–59 (2022).
15. Fei, H., Ren, Y., Zhang, Y. & Ji, D. Nonautoregressive encoder-decoder neural framework for end-to-end aspect-based sentiment triplet extraction. *IEEE Trans. Neural Netw. Learn. Syst.* **34**, 5544–5556 (2021).
16. Nie, Y., Fu, J., Zhang, Y. & Li, C. Modeling implicit variable and latent structure for aspect-based sentiment quadruple extraction. *Neurocomputing* **586**, 127642 (2024).
17. Zhang, Y., Xu, H., Zhang, D. & Xu, R. A hybrid approach to dimensional aspect-based sentiment analysis using Bert and large language models. *Electronics* **13**, 3724 (2024).
18. Manias, G., Mavrogiorgou, A., Kiourtis, A., Symvoulidis, C. & Kyriazis, D. Multilingual text categorization and sentiment analysis: A comparative analysis of the utilization of multilingual approaches for classifying twitter data. *Neural Comput. Appl.* **35**, 21415–21431 (2023).
19. Ferdosian, P. et al. Improving the efficiency and effectiveness of multilingual classification methods for sentiment analysis. In *2021 Systems and Information Engineering Design Symposium (SIEDS)*, 1–4 (IEEE, 2021).
20. Jafari, A. R., Heidary, B., Farahbakhsh, R., Salehi, M. & Crespi, N. Language models for multi-lingual tasks-a survey. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* (2024).
21. Zhang, X., Mao, R. & Cambria, E. Multilingual emotion recognition: Discovering the variations of lexical semantics between languages. In *2024 International Joint Conference on Neural Networks (IJCNN)* (2024).

22. Zhang, J., Zhao, Y., Zhang, S., Zhao, R. & Bao, S. Enhancing cross-lingual emotion detection with data augmentation and token-label mapping. In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, 528–533 (2024).
23. Teferra, B. G. et al. Screening for depression using natural language processing: Literature review. *Interact. J. Med. Res.* **13**, e55067 (2024).
24. Safdar, S. et al. Variations of emotional display rules within and across cultures: A comparison between Canada, USA, and Japan. *Can. J. Behav. Sci./Revue Canadienne des Sciences du Comportement* **41**, 1 (2009).
25. Morand, D. A. Politeness as a universal variable in cross-cultural managerial communication. *Int. J. Organ. Anal.* **4**, 52–74 (1996).
26. Latif, S., Qadir, J. & Bilal, M. Unsupervised adversarial domain adaptation for cross-lingual speech emotion recognition. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 732–737. https://doi.org/10.1109/ACII.2019.8925513 (2019).
27. Xing, H. et al. Cross-lingual word embedding generation based on Procrustes-Hungarian linear projection. In *2024 International Conference on Asian Language Processing (IALP)*, 1–6 (IEEE, 2024).
28. Nakayama, M. & Wan, Y (Sentiment balance at the aspect level for a subjective good. ScholarSpace, Cross-cultural examination on content bias and helpfulness of online reviews, 2019).
29. Gao, X., Huang, W. & Ryu, S. Effects of sentiment and emotion of campaign pitch on crowdfunding performance: A cross-cultural comparison. *J. Res. Emerg. Mark.* **3**, 23–34 (2021).
30. Jayakody, D. *et al.* Aspect-based sentiment analysis techniques: A comparative study. In *2024 Moratuwa Engineering Research Conference (MERCon)*, 205–210 (IEEE, 2024).
31. Xiang, Y., He, H. & Zheng, J. Aspect term extraction based on MFE-CRF. *Information* **9**, 198 (2018).
32. Perikos, I. & Diamantopoulos, A. Explainable aspect-based sentiment analysis using transformer models. *Big Data Cogn. Comput.* **8**, 141 (2024).
33. Chen, Y., Kong, L. & Wang, Y. Multi-grained attention representation with albert for aspect-level sentiment classification. *IEEE Access* **9**, 106703–106713 (2021).
34. Jiang, Q., Chen, L., Xu, R., Ao, X. & Yang, M. A challenge dataset and effective models for aspect-based sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference On Natural Language Processing (EMNLP-IJCNLP)*, 6280–6285 (2019).
35. Sharma, T. & Kaur, K. Aspect sentiment classification using syntactic neighbour based attention network. *J. King Saud Univ.-Comput. Inf. Sci.* **35**, 612 (2023).
36. Zheng, G., Wang, J., Yu, L.-C. & Zhang, X. Instruction tuning with retrieval-based examples ranking for aspect-based sentiment analysis. arXiv preprint arXiv:2405.18035 (2024).
37. Wang, J., Yu, L.-C. & Zhang, X. Softmcl: Soft momentum contrastive learning for fine-grained sentiment-aware pre-training. arXiv preprint arXiv:2405.01827 (2024).
38. Zhang, Y., Wang, J., Xu, D., Zhang, X. et al. YNU-HPCC at SIGHAN-2024 dimABSA task: Using PLMS with a joint learning strategy for dimensional intensity prediction. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10)*, 96–101 (2024).
39. Wang, Z., Huang, D., Cui, J. & Zhang, X. & Ho, S Subjects, methods, and trends (Sentic.net, A review of chinese sentiment analysis, 2024).
40. Fei, D. Research on interpretable text sentiment analysis. *Tokushima University Repository* (2024).
41. Hu, G. et al. Recent trends of multimodal affective computing: A survey from NLP perspective. arXiv (2024).
42. Zulqarnain, M., Ghazali, R., Aamir, M. & Hassim, Y. M. M. An efficient two-state GRU based on feature attention mechanism for sentiment analysis. *Multimed. Tools Appl.* **83**, 3085–3110 (2024).
43. Lian, Z., Tao, J., Liu, B. & Huang, J. Conversational emotion analysis via attention mechanisms. arXiv preprint arXiv:1910.11263 (2019).
44. Thakkar, G., Preradović, N. & Tadić, M. Transferring sentiment cross-lingually within and across same-family languages. *Appl. Sci.* **14**, 5652 (2024).
45. Rayhan32. Trip advisor New york city restaurants dataset (10k). https://www.kaggle.com/datasets/rayhan32/trip-advisor-newyork-city-restaurants-dataset-10k (2023). Accessed 2024 Nov 14.
46. D4rklucif3r. Restaurant reviews. https://www.kaggle.com/datasets/d4rklucif3r/restaurant-reviews (2023). Accessed: 2024 Nov 14.
47. Leone, S. Tripadvisor european restaurants dataset. https://www.kaggle.com/datasets/stefanoleone992/tripadvisor-european-restaurants/data (2023). Accessed 2024 Nov 14.

Author contributions

Arifur Rahman conceived the study, developed the methodology, and registered the initial protocol. MD Azam Khan and Kanchon Kumar Bishnu oversaw the data collection process. Arifur Rahman, MD Azam Khan and Kanchon Kumar Bishnu were responsible for analyzing the data and grading adverse events. Arifur Rahman and MD Azam Khan performed statistical analysis; Arifur Rahman, MD Azam Khan, Kanchon Kumar Bishnu and Uland Rozario prepared the initial study draft. M. F. Mridha and Zeyar Aung supervised the project. Arifur Rahman, MD Azam Khan, Kanchon Kumar Bishnu, Uland Rozario, Adit Ishraq, M. F. Mridha and Zeyar Aung contributed to manuscript preparation, critically revised the manuscript, and approved the final draft.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.