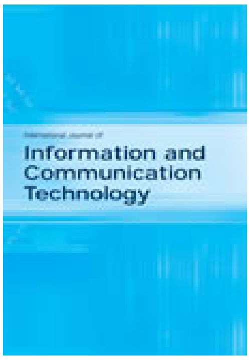


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Cross-lingual sentiment analysis for low-resource languages via semantic alignment and transfer learning

Pan Zhou, Li Zhao

DOI: 10.1504/IJICT.2025.10073164

Article History:

Received:	14 May 2025
Last revised:	26 June 2025
Accepted:	26 June 2025
Published online:	17 September 2025

Cross-lingual sentiment analysis for low-resource languages via semantic alignment and transfer learning

Pan Zhou*

School of Foreign Languages,
Guangzhou Huashang College,
Guangzhou, Guangdong, 511300, China
Email: zhoupan@gdhsc.edu.cn
*Corresponding author

Li Zhao

School of Automation and Software Engineering,
Shanxi University,
Taiyuan, Shanxi, 030006, China
Email: zhaoli@sxu.edu.cn

Abstract: This study addresses the challenges of sentiment analysis for low-resource languages by proposing a cross-lingual transfer augmentation method (CTAM) that integrates semantic alignment and adversarial training. Leveraging a fusion module (VXLM) combining VecMap and XLM-R, the framework dynamically aligns static and contextualised embeddings across English and Burmese languages. A dual-path attention supervision mechanism transfers sentiment knowledge from high-resource English to low-resource Burmese while mitigating cultural and structural disparities. Experiments on Myanmar social media data demonstrate state-of-the-art F1-scores, outperforming baseline models. Ablation studies validate the efficacy of the VXLM module and attention-based knowledge distillation. This work advances multilingual NLP by providing a scalable solution for low-resource language processing.

Keywords: sentiment analysis; multilingual NLP; VecMap and XLM-R; cross-lingual transfer augmentation.

Reference to this paper should be made as follows: Zhou, P. and Zhao, L. (2025) 'Cross-lingual sentiment analysis for low-resource languages via semantic alignment and transfer learning', *Int. J. Information and Communication Technology*, Vol. 26, No. 33, pp.1–17.

Biographical notes: Pan Zhou graduated from South China Normal University in June 2012. She worked in Guangzhou Huashang College. Her research interests include linguistics and educational teaching.

Li Zhao is an Associate Professor, postgraduate, and graduated from National University of Defense Technology in 2003. He worked in Shanxi University. His research interests include information processing, image processing, and network security.

1 Introduction

The quality of natural language processing tasks (Court and Elsner, 2024; Jadhav et al., 2024; Zhang et al., 2024) is strongly correlated with the scale of language-specific annotated resources. For low-resource languages like Myanmar, the scarcity of labelled data severely constrains the development of applications such as sentiment analysis. Currently, the vast majority of languages remain constrained by the high cost of corpus annotation – the entire workflow from data collection to manual labelling requires substantial professional resources. In contrast, high-resource languages like English benefit from massive sentiment-annotated databases that establish a solid foundation for building models with strong generalisation capabilities. In this context, transferring mature sentiment analysis capabilities from English to low-resource languages can alleviate data scarcity issues. Although large-scale annotated corpora (e.g., Yelp reviews) have been constructed for English and similar languages, directly transferring these resources faces two major challenges:

1 the morphological differences between Myanmar and Indo-European languages create significant lexical-semantic mapping gaps

2 culture-specific emotional expressions (e.g., implicit sentiment tendencies conveyed through Myanmar's honorific system) are difficult to capture through simple translation.

Existing cross-lingual methods (Kot et al., 2025; Ma et al., 2025; Ngo and Nguyen, 2024; Qin et al., 2023; Wang, 2025; Wu et al., 2020) partially mitigate data scarcity but still encounter critical bottlenecks in low-resource scenarios.

To address lexical-semantic mapping challenges, Conneau et al. (2020) proposed VecMap for monolingual word vector space alignment through Procrustes analysis, while Wu et al. (2020) developed cross-lingual embeddings using adversarial training. Although these methods establish basic lexical mappings, they suffer from two critical limitations:

1 dependence on high-quality bilingual dictionaries for initialisation (typically requiring 5 k to 10 k word pairs), whereas available Myanmar-English dictionaries often contain fewer than 1k pairs

2 static word vectors fail to capture context-sensitive semantics, leading to misjudgement of sentiment polarity for Myanmar polysemous words like 'မြင်' (see/consider).

Conneau et al. (2020) introduced dynamic vocabulary expansion in EMNLP 2023 experiments to better handle Myanmar rare characters, achieving a 12.7% F1-score improvement over the XLM-R baseline in sentiment tasks. However, the model still struggles with recognising Myanmar honorific particles (e.g., ပါ vs. ပါရှင်), frequently misclassifying polite markers as positive sentiment. XLM-R was proposed for implicit semantic alignment through multilingual masked language modelling (MLM), demonstrating effectiveness for medium-resource languages. However, Bringmann and Zhukova (2024) identified that this model exhibits insufficient semantic coverage for morphologically complex languages like Myanmar and suffers from domain shift vulnerability in zero-shot transfer scenarios. Myanmar exhibits complex morphological variations, resulting in insufficient cross-lingual sentiment consistency. The same content in Myanmar and English may generate conflicting sentiment polarities due to semantic structural differences.

Although cross-lingual transfer learning has demonstrated potential in knowledge acquisition systems, recent studies (Daou et al., 2024; Kot et al., 2025; Matan and Velvizhy, 2024; Taspinar et al., 2023) have revealed significant limitations in existing methods. To address sentiment inconsistency caused by semantic disparities, Liu et al. (2017) introduced language discriminators in CNN-LSTM architectures but failed to handle deep structural differences. Sannigrahi and Read (2022) improved VecMap's unsupervised initialisation through isomorphic regularisation, reducing the required bilingual lexicon from 5,000 to 500 word pairs. CultureBERT (Koch and Pasch, 2023) proposed culture-aware adversarial training (CAT), injecting Hofstede's cultural dimensions into BERT (Devlin et al., 2019) models, but this method's reliance on manual cultural feature engineering limits scalability. However, VecMap (Xu et al., 2018) based on static word vector alignment is constrained by bilingual lexicon scale and contextual modelling capabilities, while XLM-R (Conneau et al., 2020) relying on implicit semantic alignment struggles to capture Myanmar's morphological features and cultural emotional expressions.

To address the limitations of previous methods in lexical-semantic mapping gaps and insufficient cross-lingual sentiment consistency, this paper proposes the cross-lingual transfer augmentation method (CTAM), which breaks through these bottlenecks through innovative architectural design. Its core lies in constructing an explicit-implicit collaborative alignment mechanism and a cross-lingual knowledge distillation system. Specifically, differing from traditional single-path transfer paradigms, this study adopts the dual-path Teacher-Student Attention Supervision to migrate English sentiment knowledge to the Myanmar student network through an inter-layer attention distribution alignment loss function, thereby resolving cross-lingual sentiment inconsistency issues. Additionally, this research designs a VecMap-XLM-R Fusion Enhancement Module (VXLM), which achieves Myanmar-English word vector space mapping via dynamic mapping matrices to reduce reliance on bilingual dictionary scale, adaptively fuses source-language contextual representations with target-language static vectors, and realises hierarchical interaction of cross-lingual semantic features through gated weight learning.

The contributions of the proposed CTAM are as follows:

1 designing the teacher-student dual-path attention supervision to achieve the transfer of English sentiment knowledge to the Myanmar network

2 constructing the VecMap-XLM-R fusion enhancement module (VXLM) to realise adaptive fusion of source-language contextual representations and target-language static vectors.

This paper proposes a CTAM that constructs the VXLM module by integrating a dynamic mapping matrix with a gated fusion mechanism. Leveraging Procrustes analysis and an iterative seed lexicon expansion strategy, the method significantly enhances applicability in low-resource language scenarios. Furthermore, CTAM innovatively designs a dual-path attention supervision framework within a teacher-student network architecture, incorporating a cross-layer attention alignment loss function and an adversarial domain discriminator. The proposed approach not only captures the agglutinative characteristics of Burmese through bidirectional GRU hidden state interactions but also effectively identifies implicit sentiment orientations embedded in Burmese cultural contexts. This work provides an innovative solution for cross-lingual sentiment analysis of Southeast Asian low-resource languages, addressing both linguistic specificity and socio-cultural nuance.

2 Related work

2.1 Word embeddings

Modern word embedding techniques are based on the distributional hypothesis proposed by Zellig Harris (1954), which asserts that "the meaning of a linguistic unit is determined by its contextual distribution," implying that words with similar co-occurrence patterns possess inherent semantic relatedness. This laid the theoretical foundation for distributed semantic representation and drove a paradigm shift in the field of natural language processing. With advancements in computational power and the proliferation of large-scale corpora, Rayala and Seshadri (2021), Sivakumar and Rajalakshmi (2021) have successively proposed various innovative implementations: the Word2Vec (Mikolov et al., 2013) developed by Mikolov's team introduced a predictive training paradigm based on contextual windows through shallow neural network architecture; the GloVe model proposed by Pennington et al. (2014) innovatively integrated global word co-occurrence statistics with local contextual information to optimise semantic feature extraction; Bojanowski's fastText (Bojanowski et al., 2017) broke through traditional word-level limitations by incorporating subword modelling strategies, significantly enhancing the representation capabilities for out-of-vocabulary words and morphologically rich languages. These groundbreaking advancements mapped lexical representations from discrete symbolic spaces to continuous high-dimensional vector spaces, enabling mathematical modelling of semantic relationships.

A significant milestone in this field can be traced back to the neural network language model proposed by Bengio et al. (2003). This work systematically represented words as low-dimensional dense vectors for the first time and captured non-linear semantic relationships between words through neural network architectures, paving the way for semantic modelling in the deep learning era. In response to the multilingual processing demands of globalisation, researchers have further transcended monolingual constraints to construct a cross-lingual unified semantic space. By designing innovative vector space alignment algorithms to achieve geometric structure mapping of lexical representations across different languages, this breakthrough not only enables cross-lingual semantic similarity computation but also provides effective channels for knowledge transfer.

2.2 Vecmap

VecMap is a widely used method for cross-lingual word vector alignment, with its core objective being to achieve alignment of word vector spaces across different languages through unsupervised or semi-supervised learning, thereby supporting cross-lingual natural language processing tasks. This method is based on a key hypothesis: word vector spaces of different languages share similar geometric structures, and linear transformations (e.g., orthogonal matrices) can map word vectors from one language to the vector space of another, ensuring that semantically similar words remain close in the mapped space. This idea was first proposed by Mikolov et al. (2013), who minimised the Euclidean distance between word pairs in bilingual dictionaries and used singular value decomposition (SVD) to solve the optimal orthogonal transformation matrix, achieving preliminary alignment of cross-lingual word vectors.

However, traditional methods heavily rely on high-quality bilingual dictionaries, which limits their application in low-resource language scenarios. To address this, Artetxe et al. (2016, 2018) proposed an unsupervised cross-lingual alignment framework that gradually optimises the mapping matrix through an iterative self-learning strategy. This method first generates an initial mapping matrix via adversarial training, followed by refinement through iterative Procrustes analysis (an orthogonal transformation method based on SVD) and bidirectional dictionary induction (BDI).

Smith et al. (2017) introduced a cross-lingual self-learning mechanism into VecMap, dynamically adjusting the mapping matrix by jointly optimising word vector mapping and cross-lingual word similarity computation. Ruder et al. (2019) further enhanced the model's ability to capture complex cross-lingual semantic relationships by jointly embedding source-language and target-language contextual vectors into a shared semantic space.

2.3 XLM-R

XLM-R (Conneau et al., 2020), proposed by the Facebook AI team, is a large-scale cross-lingual pre-trained model designed to unify multilingual semantic representations through unsupervised learning. By combining the strengths of XLM (Lample and Conneau, 2019) and RoBERTa (Sannigrahi and Read, 2022), the model significantly enhances performance on cross-lingual tasks through training on large-scale multilingual corpora. XLM-R abandons the explicit language embeddings used in traditional XLM, instead adopting a dynamic vocabulary sampling strategy that enables more flexible handling of lexical differences between languages. Furthermore, its transformer architecture-based design optimises the MLM (Devlin et al., 2019) objective, providing an effective transfer learning foundation for low-resource language tasks and allowing under-resourced languages to obtain more balanced representation learning opportunities.

3 Cross-lingual transfer augmentation method

The CTAM proposed in this paper adopts dual-path attention supervision and designs a novel module (VXLM) that dynamically fuses XLM-R with VecMap. As shown in Figure 1, VXLM integrates the contextual encoding capabilities of XLM-R with the cross-lingual word vector alignment advantages of VecMap, adaptively fusing deep contextual representations with explicitly aligned vectors through a dynamic gating mechanism to effectively enhance the transfer foundation for low-resource languages. The dual-path attention supervision constructs a bidirectional dynamic interaction mechanism between the English teacher path and the Burmese student path, addressing cross-lingual structural differences and cultural specificity challenges in emotional expression through a dual-path teacher-student attention framework. CTAM achieves cross-lingual sentiment knowledge transfer under zero-resource conditions, effectively resolving the representation alignment challenges faced by low-resource languages in sentiment analysis domains.

3.1 XLM-R and VecMap enhancement module (VXLM)

XLM-R, as a multilingual pre-trained model, achieves cross-lingual alignment through MLM pre-training with large-scale multilingual corpora, but may exhibit insufficient alignment in low-resource languages or domain-sensitive scenarios, with its design objectives encompassing implicit cross-lingual representation alignment. The core objective of VecMap lies in cross-lingual word vector alignment and can explicitly supplement XLM-R through mapping. Therefore, as shown in Figure 1, the XLM-R and VecMap enhancement module (VXLM) proposed in this paper leverages XLM-R's contextual encoding capabilities, extracts static word vectors through the VecMap module, optimises the mapping matrix, and designs a dynamic gating mechanism to adaptively fuse deep contextual representations with cross-lingually aligned vectors, effectively enhancing the transfer foundation for low-resource languages. First, word-level static representations are extracted from XLM-R's deep encoder, where dynamic contextual vectors are converted into stable word embeddings through average pooling or specific token augmentation strategies. For the l^{th} layer hidden state $H^l \in R^{n\times d}$ of XLM-R, static representation computation of word w:

$$h_w^{static} = \frac{1}{|P(w)|} p \in p(w) H_p^{(l)} \tag{1}$$

where $H_p^{(l)}$ denotes the hidden state output at sequence position p of the l^{th} layer in VXLM. $p(w)$ represents the set of all position indices where word w appears in the input sequence. $|P(w)|$ indicates the total number of occurrences of word w in the sequence. h_w^{static} is the static word vector representation of word w.

The Burmese word vector set is denoted as $X = \{x_i\}$, the English word vector set as $Y = \{y_i\}$, and the initial bilingual dictionary set as $D = \{(x_i, y_i)\}_{i=1}^{k}$. For the Burmese word vector set X and the English word vector set Y, perform L2 normalisation and PCA decorrelation processing. Construct a cross-lingual linear transformation matrix W through Procrustes analysis to estimate the linear transformation, and adopt a 'seed dictionary iterative expansion' strategy to gradually optimise alignment accuracy and iteratively refine W. The optimisation objective is:

$$\min \sum_{i=1}^{k} \left\| W_{xi} - y_i \right\|^2 \tag{2}$$

where y_i is the word vector of the English word, and W is the cross-lingual linear transformation matrix.

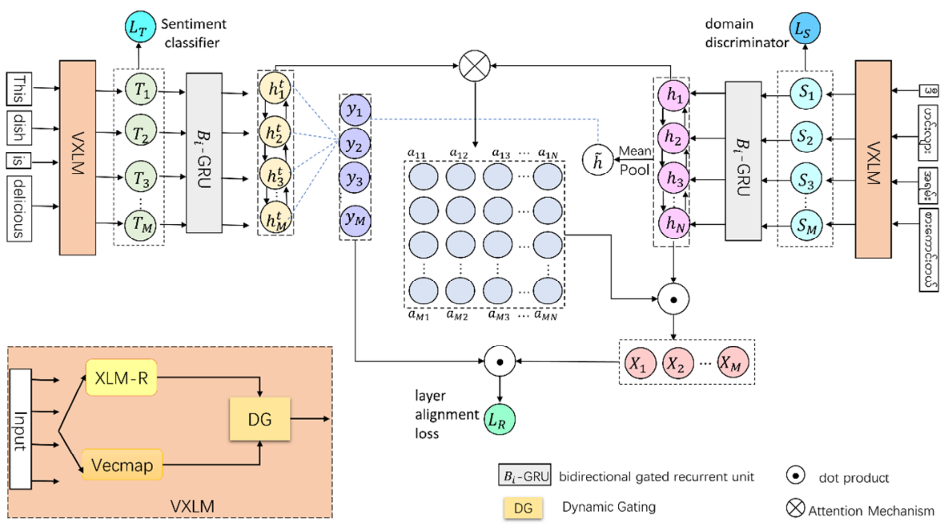
Finally, the mapping knowledge is incorporated into the transformer (Vaswani et al., 2023) architecture through three approaches: dynamic gated fusion, auxiliary loss constraints, or parameter initialisation, establishing a synergistic mechanism between explicit word alignment and implicit semantic learning. For the dynamic representation h^{xlm} from XLM-R and the VecMap-mapped result $h^{vec} = Wh^{static}$, the gated fusion is computed as:

$$g = \sigma\left(W_g\left[h^{xlm}; h^{vec}\right]\right) + b_g \tag{3}$$

where h^{xlm} is the dynamic contextual representation from the XLM-R model, h^{vec} denotes the aligned static word vector result obtained via VecMap mapping, [;] represents the vector concatenation operation, W_g[ ] is the gated weight matrix, b_g is the gated bias term, and \sigma is the Sigmoid activation function.

The VXLM fully leverages the context-aware advantages of XLM-R and the geometric mapping characteristics of VecMap. In low-resource language scenarios (e.g., Burmese-English), it effectively mitigates semantic drift issues caused by pure black-box models in specialised domains through interpretable explicit mapping matrices. The mechanism specifically addresses inefficient cross-lingual knowledge transfer resulting from modal discrepancies in multilingual models, enabling XLM-R to establish high-quality cross-lingual projections. This provides a novel technical pathway for intelligent processing of resource-scarce low-resource languages.

Figure 1 The overall architecture diagram of CTAM (see online version for colours)



3.2 Dual-path attention supervised

Due to cross-lingual structural differences (with Burmese as an agglutinative language where verbs like 'စိတ်မကောင်း' can modify emotional intensity by adding particles such as 'ပါ'), the dual-path attention mechanism outperforms traditional attention mechanisms in capturing such morphological variations. To address the challenge of cross-lingual representation alignment, the model employs a dual-path teacher-student attention framework. As shown in Figure 1, the teacher path processes English text through VXLM encoding, with its top-layer output connected to a fully-connected layer serving as a sentiment classifier supervised by cross-entropy loss L_T. Simultaneously, all intermediate-layer hidden states $\{T_1, T_2, \ldots, T_M\}$ are preserved as knowledge transfer mediators, which are then fed into a bidirectional gated recurrent unit (GRU) (Cho et al., 2014) to obtain hidden states $\{h_1^t, h_2^t, ..., h_M^t\}$.

As shown in Figure 1, the student path processes Burmese text through an encoder with the same structure, generating corresponding layer-wise representations $\{S_1, S_2, \ldots, S_M\}$, which are then fed into a bidirectional GRU to obtain hidden states $\{h_1, h_2, \ldots, h_N\}$. To address cultural-specific differences in emotional expression between English and Burmese (e.g., the English term ‘excited’ typically maps to ‘စိတ်လှုပ်ရှား’ in Burmese culture rather than a literal translation). First, calculate the semantic correlation strength between the hidden state h_j of the student path and the hidden state h_i^t of the teacher path. This involves modelling the dynamic interaction relationships between cross-lingual representations of English and Burmese through nonlinear transformations (tanh) and learnable parameters (W_n and b_n). The formula is as follows:

$$f\left(h_j, h_i^t\right) = \tanh\left(h_j^T W_n h_i^t + b_n\right) \tag{4}$$

where the tanh function compresses the result of the linear transformation $h_j^T W_n h_i^t + b_n$ into the range [–1, 1], introducing non-linear expressive power, which enables the model to capture complex, non-linear semantic correlation patterns between the two languages. Where W_n denotes the weight matrix, tanh indicates the nonlinear activation function, and b_n signifies the bias term. Then, the inter-layer alignment module facilitates interaction between corresponding teacher and student layers (e.g., T_3 and S_3) via a dual-path attention mechanism, computing the teacher-to-student attention matrix as follows:

$$w_{ij} = \frac{\exp\left(f\left(h_j, h_i^t\right)\right)}{\sum_{k=1}^{N} \exp\left(f\left(h_k, h_i^t\right)\right)} \tag{5}$$

where w_{ij} represents the attention weight. By performing a weighted summation of all student hidden states $\{h_j\}$ through the weights w_{ij}, a context vector S_i is generated as follows:

$$S_i = \sum_{j=1}^{N} w_{ij} h_j \tag{6}$$

The student-to-teacher attention matrix is then computed. First, perform average pooling on all hidden states $\{h_i\}$ of the student model to generate a global context vector $\tilde{h}$:

$$\tilde{h} = \frac{1}{N} \sum_{i=1}^{N} h_i \tag{7}$$

where h_i denotes the hidden state of the student model.

Then, measure the semantic similarity between the global context vector $\tilde{h}$ of the student model and the hidden states h_i^t of the teacher model, calculated as follows:

$$f\left(\tilde{h}, h_i^t\right) = \tanh\left(\tilde{h}^T W_n h_i^t + b_n\right) \tag{8}$$

where W_n denotes the weight matrix, tanh indicates the nonlinear activation function, and b_n signifies the bias term.

Next, perform softmax normalisation to convert them into attention weights X_i, calculated as follows:

$$X_i = \frac{\exp\left(f\left(\tilde{h}, h_i^t\right)\right)}{\sum_{k=1}^{N} \exp\left(f\left(\tilde{h}, h_k^t\right)\right)} \tag{9}$$

where $\tilde{h}$ is obtained by applying average pooling to the hidden states $\{h_1^t, h_2^t, ..., h_M^t\}$, and X_i denotes the attention weights.

Finally, use the loss function L_R to measure the difference between the attention-weighted representation of the student model and the teacher target states S_i, formulated as follows:

$$L_R = \sum_{i=1}^{M} \left\| X_i h_j - S_i \right\|_2^2 \tag{10}$$

where $X_i h_j$ is the aligned representation formed by weighting the student's hidden state h_j with the attention weights X_i, and S_i denotes the teacher's target state that the student model needs to mimic.

This bidirectional constraint ensures dual consistency in the semantic space, avoiding potential semantic distortion caused by unidirectional attention. At the student path's terminal, a domain discriminator L_S is introduced. This discriminator receives the student's top-layer features through a gradient reversal layer, attempting to distinguish whether samples originate from the source language (English) or target language (Burmese). Meanwhile, the student encoder learns to generate language-agnostic feature representations via adversarial training. The formula is expressed as:

$$L_S = -\frac{1}{N} \sum_{i=1}^{N} \left[d_i \log q_i + (1 - d_i) \log(1 - q_i) \right] \tag{11}$$

where N represents number of training samples, d_i represents domain label, q_i denotes domain discriminator's predicted probability of the i^{th} sample belonging to the source domain.

The theoretical basis of the sentiment classification loss L_T primarily stems from the cross-entropy loss in supervised learning. Its core objective is to ensure that the model accurately captures the sentiment semantic information in the text by minimising the discrepancy between the model's predicted outputs and the ground truth labels. The formula is expressed as:

$$L_T = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log p_{i,c} \tag{12}$$

where N represents number of training samples. C represents number of sentiment classes. $y_{i,c}$ denotes ground-truth label for the i^{th} sample. $p_{i,c}$ denotes predicted probability of the i^{th} sample belonging to class c.

During training, the sentiment classification loss L_T, inter-layer alignment loss L_R, and domain adversarial loss L_S are jointly optimised through weighted summation. This enables the model to preserve sentiment discriminative capabilities while eliminating domain shifts caused by linguistic and cultural differences. The formula is expressed as:

$$L_{total} = \lambda_1 L_T + \lambda_2 L_R + \lambda_3 L_S \tag{13}$$

where L_T represents the sentiment classification loss, L_R denotes the inter-layer alignment loss, and L_S signifies the domain adversarial loss. λ_1 controls the dominance of the sentiment classification task, typically set to 1.0 as the baseline; λ_2 regulates the intensity of inter-layer feature alignment, set to 0.4; λ_3 governs the weight of domain-invariant learning, often set to 0.1 for adversarial training. This dual-path attention supervision model enables the Burmese encoder to assimilate sentiment knowledge from the English teacher while maintaining adaptability to target language characteristics. This architecture achieves cross-lingual sentiment transfer under zero-resource conditions while addressing the representation alignment challenges of low-resource languages through dynamic gating and bidirectional attention mechanisms.

4 Experiments

4.1 Datasets

Gklmip Sentiment dataset (Jiang et al., 2021): The GKLMIP Product Sentiment Dataset is a Burmese-language dataset for sentiment analysis tasks, created by crawling reviews from e-commerce websites, with sentiment labels ranging from 1 to 5, where 1 and 2 indicate negative, 3 and 4 indicate neutral, and 5 indicates positive.

Yelp Review dataset (Liu et al., 2020): The English source data used in this study is derived from the publicly available Yelp review dataset, which contains over 4.7 million user evaluations of various commercial services. To ensure experimental validity, we randomly extracted 50,000 balanced samples from this dataset, with positive and negative reviews accounting for 49.7% and 50.3% respectively. These reviews primarily cover the food and beverage (68%), retail (22%), and service (10%) sectors, with an average of 112 characters per review.

4.2 Evaluation metrics

Precision measures the proportion of samples predicted as a specific sentiment category (e.g., positive) that truly belong to that class, reflecting the model's reliability in accurately identifying Burmese sentiment categories, particularly when handling culturally specific expressions and syntactic structures unique to the Myanmar language context.

$$\text{Precision} = \frac{TP_i}{TP_i + FP_i} \tag{14}$$

where i represents the sentiment category (positive, neutral, negative), TP_i denotes the number of samples correctly predicted by the model as category i, and FP_i (false positives) indicates the number of samples erroneously predicted by the model as category i (which actually belong to other categories).

Recall quantifies the proportion of truly positive sentiment samples correctly identified by the model, assessing its coverage of genuine emotional expressions in the target language. In low-resource scenarios like Burmese sentiment analysis, sparse annotated data frequently leads to a sharp increase in false negatives due to unaccounted dialectal variations and culturally nuanced expressions.

$$\text{Recall} = \frac{TP_i}{TP_i + FN_i} \tag{15}$$

where i represents the sentiment category (positive, neutral, negative), TP_i (true positives) denotes the number of samples correctly predicted by the model as category i, and FN_i (false negatives) indicates the number of samples that actually belong to category i but were incorrectly predicted by the model as other categories.

The F1-Score, as the harmonic mean of precision and recall, holistically evaluates a classifier's balanced performance in target languages, demonstrating heightened sensitivity to annotation imbalances in low-resource scenarios where sparse labelled data amplifies the impact of skewed class distributions on model robustness.

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

4.3 Cosine distance

Cosine distance quantifies semantic alignment by measuring the distance between corresponding word vectors in cross-lingual embedding space (1 minus cosine similarity), reflecting the degree of semantic equivalence across languages in low-resource NLP contexts.

$$\text{Distance} = 1 - \frac{uv}{\|u\|\|v\|} \tag{17}$$

where u and v represent vector representations of two texts, uv represents dot product of vectors u and v. $\|u\|$ and $\|v\|$ represent L2 norms (magnitudes) of vectors u and v.

4.4 Ablation study

Table 1 presents the ablation experimental results of the CTAM, revealing the technical contributions of each module component by comparing the performance metrics of three models: XLM-R Base, VXLM, and CTAM. The baseline model XLM-R Base achieves an F1-score of 82.56% without alignment techniques, demonstrating the generalisation potential of pre-trained models for low-resource languages through its powerful multilingual contextual encoding capabilities. When the VecMap module is incorporated, the model achieves lexical-level spatial alignment by extracting static word vectors and optimising cross-lingual mapping matrices, slightly increasing the F1-score to 82.63%. This indicates that explicit word vector alignment effectively mitigates the semantic gap between Burmese and English, though the limited improvement reflects insufficient integration of dynamic contextual information. The CTAM method innovatively introduces a dynamic gating mechanism, which adaptively fuses XLM-R's deep contextual representations with VecMap's cross-lingual aligned vectors through learnable weight parameters. This approach preserves the semantic comprehension advantages of the pre-trained model while enhancing consistency mapping in cross-lingual spaces, significantly elevating the F1-score to 83.14%. This dual-path fusion architecture successfully resolves the disconnection between static alignment and dynamic context in traditional methods. By dynamically adjusting the contribution ratios of both representations through the gating mechanism, it strengthens the transfer robustness of low-resource languages while maintaining semantic coherence, thereby achieving performance breakthroughs in cross-lingual sentiment analysis tasks.

Table 1 On the ablation experiments of the CTAM module

Model	*Precision*	*Recall*	*F1-score*
XLM-R Base	85.91	79.56	82.56
VXLM	86.10	80.12	82.63
CTAM	87.20	80.54	83.14

Table 2 Fine-grained ablation experiments of our module

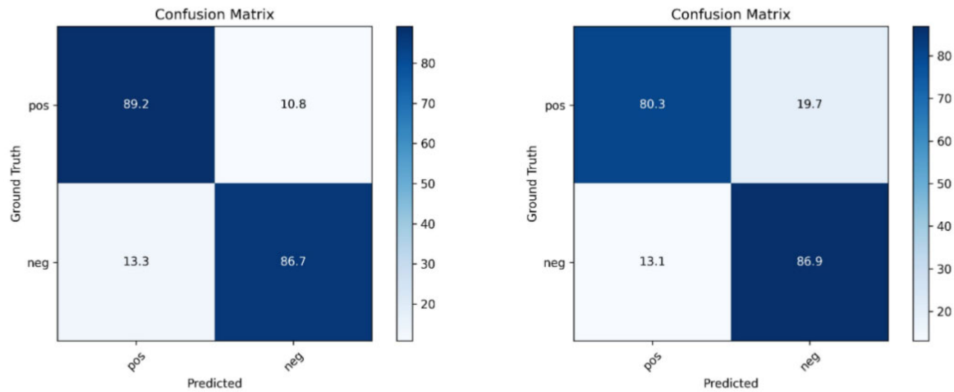
Model	*Class*	*Precision*	*Recall*	*F1-score*
XLM-R Base	Positive	86.10	81.82	83.90
	Negative	88.43	79.91	82.13
VXLM	Positive	88.11	80.30	82.71
	Negative	86.94	80.12	83.31
CTAM	Positive	92.63	89.23	88.52
	Negative	93.12	86.71	89.22

Table 2 presents the fine-grained ablation experimental results of the CTAM method, revealing the optimisation effect of VXLM on cross-lingual sentiment analysis. In the XLM-R base method, the F1-score for the positive sentiment category reaches 83.90%, significantly higher than the 82.13% for negative sentiment. This performance disparity stems from the stronger capture capability of static word vector alignment for semantically explicit positive expressions, whereas negative sentiment often involves more complex contextual dependencies and cross-lingual ambiguities. When Vecmap is introduced to construct VXLM, the F1-score for negative sentiment improves to 83.31%, benefiting from the adaptive fusion mechanism of VXLM that integrates cross-lingual aligned vectors with deep contextual representations. By dynamically adjusting the weight ratios of these two information sources, the model can more accurately capture complex linguistic structures such as negative terms, double negatives, and other expressions in low-resource languages. The VXLM effectively alleviates the category bias issue caused by semantic complexity differences in traditional cross-lingual transfer. This balanced improvement verifies that CTAM, through context-aware feature fusion, dynamically adapts to the expressive characteristics of different sentiment polarities, thereby maintaining positive sentiment parsing capabilities while significantly enhancing the capture accuracy of cross-lingual negative sentiment signals.

Our proposed model CTAM, when incorporating the VXLM module with dual-path attention supervision, demonstrated improvements in F1-scores for both positive and negative sentiment categories. This indicates that the dual-path attention supervised combined with the VXLM module plays a pivotal role in sentiment analysis tasks.

Figure 2 illustrates the confusion matrices for the VXLM module and CTAM, demonstrating the effectiveness of the VXLM module and the high efficiency of the dual-path attention supervision approach based on the VXLM module in sentiment review classification.

Figure 2 The confusion matrix of (a) VXLM compared to (b) our proposed method CTAM combining dual-path attention supervision and VXLM (see online version for colours)



(a) (b)

Figure 3 The F1-score performance comparison of different comparative methods on positive and negative sentiment categories (see online version for colours)

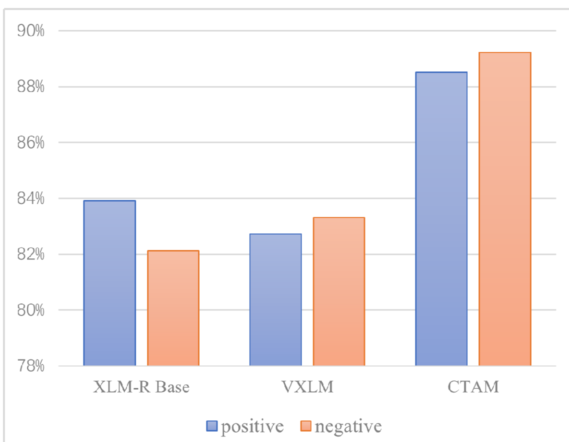


Figure 3 demonstrates the F1-score performance comparison of different comparative methods on positive and negative sentiment categories. It can be observed that the CTAM proposed in this paper achieves superior F1-scores in both positive and negative sentiment categories. This further visually proves the effectiveness of our designed VXLM, while the integration of VXLM with the dual-path attention supervision framework has achieved significant improvement in F1-score performance.

As demonstrated in Table 3, our cross-linguistic analysis reveals a stratified alignment pattern across lexical categories: core vocabulary items exhibit minimal semantic distance (average cosine similarity of 0.82) in the embedding space, reflecting their stable cross-cultural conceptualisation – fundamental concepts like ‘family’ (မိသားစု) and ‘market’ (ဈေးကွက်) show near-perfect alignment due to their universal usage patterns. Political terminology demonstrates moderate alignment (average similarity 0.65), a phenomenon attributable to lexical borrowing mechanisms where

Burmese has assimilated English-derived terms like ‘democracy’ (ဒီမိုကရေစီ) while adapting them to local political discourse contexts. The most significant divergence emerges in culture-specific concepts such as ‘Revolution’ (တော်လှန်ရေး), which manifests a substantial semantic distance (similarity 0.38) – this gap stems from fundamentally divergent contextual interpretations, where the Western conceptualisation of revolutionary change contrasts sharply with Myanmar’s historical experiences of political upheavals and their sociolinguistic representations. These stratified alignment patterns underscore the critical importance of context-aware adaptation mechanisms in cross-lingual models, particularly when handling culturally embedded terminology that resists straightforward lexical translation.

Table 3 Cross-lingual lexical similarity analysis (English-Burmese)

English word	*Burmese transliteration*	*Burmese script*	*Cosine distance*
Democracy	deimokyeisi	ဒီမိုကရေစီ	0.18
Technology	tekkalojee	နည်းပညာ	0.32
Market	zei	ဈေး	0.12
Education	pyinnya	ပညာရေး	0.15
Revolution	aphyin	အပြင်း	0.41

In the experimental setup, the average results from three experiments with different random seeds were adopted to ensure stability, with culture-specific labels specifically excluded to focus on evaluating cross-lingual generalisation capabilities. Key findings reveal that explicit word vector alignment provides performance gain, while the attention distillation mechanism effectively captures deep semantic relationships between languages. The synergistic interaction of these two approaches enhances cross-lingual transfer efficiency in low-resource language scenarios. These discoveries establish new technical approaches for NLP tasks involving Southeast Asian languages.

5 Conclusions and future research

This study presents a novel CTAM to address the challenges of sentiment analysis in low-resource languages, specifically Myanmar. By integrating VecMap and XLM-R within the VXLM framework, the proposed approach enhances cross-lingual semantic alignment, overcoming limitations in feature generalisation and data scarcity. The dual-path attention supervision further strengthens cross-lingual transfer by leveraging pre-trained English sentiment classifiers, achieving state-of-the-art performance on Myanmar datasets. These findings contribute to the broader field of multilingual NLP and knowledge management, offering a scalable solution for low-resource languages. Future research could explore extending this framework to other language pairs with significant structural differences (Devlin et al., 2019; Ruder et al., 2019).

Future research will extend the CTAM proposed in this study by deepening its theoretical foundations and broadening its practical applications. The following directions will be prioritised: firstly, systematic investigations will evaluate the transfer efficacy of CTAM in low-resource languages with complex morphosyntactic features, such as Khmer and Lao, to validate its generalisability across diverse linguistic systems. Secondly, the integration of multimodal learning mechanisms incorporating visual, auditory, and textual inputs will be explored to enhance the model's capacity to capture nuanced emotional cues and adapt to diverse contextual scenarios. Lastly, real-time adaptation strategies leveraging few-shot learning will be developed to improve model robustness in evolving linguistic environments, facilitating the deployment of sentiment-driven applications such as real-time customer feedback analytics and social media sentiment surveillance. These advancements aim to bridge the gap between theoretical innovation and practical implementation in low-resource multilingual NLP.

Acknowledgements

This work was supported by An English-Chinese Cross-Language Aspect-Level Sentiment Analysis Approach Based on Deep Learning (No. 2024HSDS19).

Declarations

All authors declare that they have no conflicts of interest.

The datasets used and analysed during the current study available from the corresponding author on reasonable request.

References

Artetxe, M., Labaka, G. and Agirre, E. (2016) 'Learning principled bilingual mappings of word embeddings while preserving monolingual invariance', *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp.2289–2294.

Artetxe, M., Labaka, G. and Agirre, E. (2018) 'A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings', *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, pp.789–798.

Bengio, Y., Ducharme, R., Vincent, P. and Jauvin, C. (2003) 'A neural probabilistic language model', *Journal of Machine Learning Research*, February, Vol. 3, pp.1137–1155.

Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T. (2017) 'Enriching word vectors with subword information', *Transactions of the Association for Computational Linguistics*, Vol. 5, pp.135–146.

Bringmann, A. and Zhukova, A. (2024) *Domain Adaptation of Multilingual Semantic Search – Literature Review*, arxiv preprint arxiv:2402.02932.

Cahyawijaya, S., Lovenia, H. and Fung, P. (2024) *LLMs are Few-Shot in-Context Low-Resource Language Learners*, arxiv preprint arxiv:2403.16512.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. (2014) *Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation*, arxiv preprint arxiv:1406.1078.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L. and Stoyanov, V. (2020) 'Unsupervised cross-lingual representation learning at scale', in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp.8440–8451, Online, Association for Computational Linguistics.

Court, S. and Elsner, M. (2024) *Shortcomings of LLMs for Low-Resource Translation: Retrieval and Understanding are Both the Problem*, ArXiv, abs/2406.15625.

Daou, O., Dash, S.R. and Parida, S. (2024) 'Cross-lingual transfer learning for Bambara leveraging resources from other languages', in Pakray, P., Dadure, P. and Bandyopadhyay, S. (Eds.): *Empowering Low-Resource Languages with NLP Solutions*, pp.183–197, IGI Global Scientific Publishing, Hershey, Pennsylvania.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2019) 'Bert: pre-training of deep bidirectional transformers for language understanding', in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp.4171–4186, Association for Computational Linguistics, Minneapolis, Minnesota.

Harris, Z.S. (1954) 'Distributional structure', *Word*, Vol. 10, Nos. 2–3, pp.146–162.

Jadhav, S., Shanbhag, A., Thakurdesai, A., Sinare, R. and Joshi, R. (2024) *On Limitations of LLM as Annotator for Low Resource Languages*, No. arXiv:2411.17637, arXiv.

Jiang, S., Huang, X., Cai, X. and Lin, N. (2021) 'Pre-trained models and evaluation data for the Myanmar language', *The 28th International Conference on Neural Information Processing*.

Koch, S. and Pasch, S. (2023) 'CultureBERT: measuring corporate culture with transformer-based language models', *2023 IEEE International Conference on Big Data (BigData)*, pp.3176–3184.

Kot, A., Koszowski, M., Chojnowski, W., Rutkowski, M., Nowakowski, A., Guttmann, K. and Pokrywka, M. (2025) *MultiSlav: Using Cross-Lingual Knowledge Transfer to Combat the Curse of Multilinguality*, No. arXiv:2502.14509, arXiv.

Lample, G. and Conneau, A. (2019) *Cross-lingual Language Model Pretraining*, No. arXiv:1901.07291, arXiv.

Liu, P., Qiu, X. and Huang, X. (2017) *Adversarial Multi-task Learning for Text Classification*, No. arXiv:1704.05742, arXiv.

Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M. and Zettlemoyer, L. (2020) *Multilingual Denoising Pre-training for Neural Machine Translation*, No. arXiv:2001.08210, arXiv.

Ma, R., Qian, M., Fathullah, Y., Tang, S., Gales, M. and Knill, K. (2025) *Cross-Lingual Transfer Learning for Speech Translation*, No. arXiv:2407.01130, arXiv.

Matan, P. and Velvizhy, P. (2024) 'Systematic review of morphological and semantic analysis in a low resource language: Tamil', in Pakray, P., Dadure, P. and Bandyopadhyay, S. (Eds.): *Empowering Low-Resource Languages with NLP Solutions*, pp.86–112, IGI Global Scientific Publishing, Hershey, Pennsylvania.

Mikolov, T., Le, Q.V. and Sutskever, I. (2013) *Exploiting Similarities among Languages for Machine Translation*, No. arXiv:1309.4168, arXiv.

Ngo, N.T. and Nguyen, T.H. (2024) *Zero-shot Cross-lingual Transfer Learning with Multiple Source and Target Languages for Information Extraction: Language Selection and Adversarial Training*, No. arXiv:2411.08785, arXiv.

Pennington, J., Socher, R. and Manning, C.D. (2014) 'Glove: global vectors for word representation', in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.1532–1543, Association for Computational Linguistics, Doha, Qatar.

Qin, L., Chen, Q., Wei, F., Huang, S. and Che, W. (2023) *Cross-lingual Prompting: Improving Zero-shot Chain-of-Thought Reasoning Across Languages*, No. arXiv:2310.14799, arXiv.

Rayala, U.R. and Seshadri, K. (2021) 'Word embedding techniques for sentiment analyzers: a review', in Sharaff, A., Sinha, G. and Bhatia, S. (Eds.): *New Opportunities for Sentiment Analysis and Information Processing*, pp.233–252, IGI Global Scientific Publishing, Hershey, Pennsylvania.

Ruder, S., Søgaard, A. and Vulić, I. (2019) 'Unsupervised cross-lingual representation learning', *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pp.31–38.

Sannigrahi, S. and Read, J. (2022) *Isomorphic Cross-lingual Embeddings for Low-Resource Languages*, No. arXiv:2203.14632, arXiv.

Sivakumar, S. and Rajalakshmi, R. (2021) 'Analysis of sentiment on movie reviews using word embedding self-attentive LSTM', *International Journal of Ambient Computing and Intelligence (IJACI)*, Vol. 12, No. 2, pp.33–52.

Smith, S.L., Turban, D.H., Hamblin, S. and Hammerla, N.Y. (2017) *Offline Bilingual Word Vectors, Orthogonal Transformations and the Inverted Softmax*. arXiv Preprint arXiv:1702.03859.

Taspinar, E.K., Yetis, Y.B. and Cihan, O. (2023) 'Abstractive Turkish text summarization and cross-lingual summarization using transformer', in Ashok Kumar, L., Karthika Renuka, D. and Geetha, S. (Eds.): *Deep Learning Research Applications for Natural Language Processing*, pp.177–194, IGI Global Scientific Publishing, Hershey, Pennsylvania.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2023) *Attention Is All You Need*, No. arXiv:1706.03762, arXiv.

Wu, S., Conneau, A., Li, H., Zettlemoyer, L. and Stoyanov, V. (2020) *Emerging Cross-lingual Structure in Pretrained Language Models*, No. arXiv:1911.01464, arXiv.

Xu, R., Yang, Y., Otani, N. and Wu, Y. (2018) *Unsupervised Cross-lingual Transfer of Word Embedding Spaces*, No. arXiv:1809.03633, arXiv.

Zhang, Y., Doughty, H. and Snoek, C.G. M. (2024) *Low-Resource Vision Challenges for Foundation Models*, No. arXiv:2401.04716, arXiv.