

# A DOMAIN-INVARIANT TRANSFER LEARNING BY BERT FOR CROSS-DOMAIN SENTIMENT ANALYSIS

Al-Mahmud and Kazutaka Shimada

Department of Artificial Intelligence, Kyushu Institute of Technology, Fukuoka, Japan

## ABSTRACT

*Sentiment analysis is aimed at analyzing the attitudes or behaviors of individuals and entities from user-generated content. Cross-domain sentiment analysis is the process of examining sentiments expressed in textual data across diverse domains or topics. Unlike traditional sentiment analysis, which concentrates on a particular domain or topic, cross-domain sentiment analysis entails transferring knowledge from trained models on one domain to another domain. It is effective where the labeled data are scarce or unavailable. In this study, our target language is the Bangla language. Numerous traditional machine-learning-based sentiment analysis approaches have been proposed in the Bangla language. They often require a large amount of data to build robust models. However, manual collection/annotation of much training data within the same domain (i.e., domain-specific) can be costly, especially in low-resource languages like Bangla. To address this challenge, we collect publicly available data in one source domain (e.g., drama) by exploiting auxiliary information from it to assist the target domain (e.g., cricket) data/task. Then the model is re-trained and evaluated on the target domain (e.g., cricket) data. We establish various baselines using machine-learning-based and transformer-based models. The baselines are unable to reduce the domain gap between the source and target domains. To this end, we propose a domain-invariant transfer learning approach to bridge the domain gap. We conduct experiments and make comparative analyses between our proposed approach and the baselines. The experimental results demonstrate that the proposed approach outperforms all the baselines and exhibits its efficacy.*

## KEYWORDS

*Sentiment analysis, Cross-domain sentiment analysis, Source data, Target data, Machine-learning-based models, Transformers-based models, Combined data approach, Stepwise learning approach & Domain-invariant transfer learning approach.*

## 1. INTRODUCTION

In today's information age, people express their opinions on online platforms, including popular ones like Facebook, Twitter, and Amazon. This trend has led to an exponential increase in the volume of opinion content on the Internet [1]. Sentiment analysis, a field of study within natural language processing (NLP), focuses on analyzing individuals' subjective attitudes. It encompasses a range of sentiments such as like or dislike, preference or aversion, directed towards entities such as products, topics, or issues, as well as specific aspects like price or quality [2]. Now, let's examine three instances of reviews within the *restaurant* domain.

e1) Burgers are awesome.

Here, the writer expressed a positive opinion about burgers or food.

e2) Waiters are ordinary.

In e2, the writer expressed a negative opinion about the waiters or service.

e3) Ambiance is so-so.

In e3, the writer expressed a neutral opinion about the ambiance or atmosphere of the restaurant.

Cross-domain sentiment analysis refers to the task of analyzing sentiments expressed in textual data across different domains or topics. Unlike conventional sentiment analysis, which concentrates on a particular domain or topic, cross-domain sentiment analysis entails transferring knowledge from trained models on one domain to another domain. It is fruitful where the labeled data are inadequate or unavailable. This task is particularly challenging due to differences in vocabulary use, and sentiment expressions across different domains. The goal of cross-domain sentiment analysis is to develop robust models that can effectively generalize across different domains, thereby reducing the need for domain-specific labeled data and improving the applicability of sentiment analysis in real-world scenarios [3-5].

Sentiment analysis plays a crucial role in the realm of e-commerce. For retailers, analyzing customer-generated reviews is paramount in understanding the strengths and weaknesses of their products or services. This analytical insight serves as a foundation for refining policies and strategies, ultimately giving retailers a competitive advantage in the market.

In this study, our target language is Bangla. Various traditional machine-learning-based sentiment analysis approaches have been proposed in the Bangla language. They often require a large amount of data to generate vigorous models. However, manual collection/annotation of much training data within the same domain (i.e., domain-specific) can be expensive, especially in low-resource languages like Bangla. Despite Bangla being the seventh most spoken language globally, with approximately 265 million native and non-native speakers, gathering sufficient training data poses a significant challenge [6] due to the limited availability of annotated datasets in a particular domain [7]. To navigate this challenge, we concentrate on cross-domain sentiment analysis in the Bangla language. In cross-domain sentiment analysis, we collect a Bangla sentiment analysis dataset from the *drama*<sup>1</sup> domain which is publicly available. In addition, we obtain Bangla sentiment analysis datasets in the *cricket* and *restaurant* domain from a publicly available source<sup>2</sup>. Therefore, there are six combinations of these three distinct datasets: *drama & cricket*, *drama & restaurant*, *cricket & drama*, *restaurant & drama*, *cricket & restaurant*, and *restaurant & cricket*. The left side of the symbol “&” is treated as the “sourcedata” and the right side is treated as the “target data”. We utilize all six combinations in the experiment.

The existing studies [3-5] performed cross-domain sentiment analysis by utilizing contrastive learning, adversarial and domain-aware learning, and adversarial soft prompt tuning techniques, respectively. They performed the task in the English language with relatively much data. These methods are robust but relatively complex due to their model architectures and training procedures. However, we propose a relatively simple yet effective approach, namely domain-invariant transfer learning in the Bangla language with relatively small data. The proposed approach consists of two stages. For the first stage, we fine-tune/train only the last encoder layer of BERT with the merged source, and training and validation sets of the target datasets to bridge the domain gap between the source and target data. For the second stage, we re-use the updated model obtained from the first stage and re-train and evaluate the model utilizing the target data. Note that in the second stage, we only fine-tune/train the linear layers on top of BERT but not the BERT encoder layers (the detailed procedure is illustrated in Section 4.3).

Our study makes several key contributions, which can be summarized as follows:

1. We established five baselines leveraging machine-learning-based and transformers-based models.
2. We performed an exhaustive evaluation across domains by the combination of different datasets.

<sup>1</sup> <https://github.com/sazzadcse/BN-Dataset>

<sup>2</sup> [https://github.com/atik-05/Bangla\\_ABSA\\_Datasets](https://github.com/atik-05/Bangla_ABSA_Datasets)

- We compared the baseline performance with our proposed approach and demonstrated the effectiveness of the proposed approach.

## 2. DATASET

The sentiment data for the *cricket* and *restaurant* domain in the Bangla language is collected from a publicly available source<sup>2</sup>. The *cricket* dataset comprises 2,979 instances (2,152 negative, 566 positive, and 261 neutral) whereas the restaurant dataset contains 2,059 instances (472 negative, 366 positive, and 1,221 neutral). Detailed discussions on these datasets can be found in the referenced study [8].

Again, we accumulated the sentiment data for the *drama* domain from another publicly available source<sup>1</sup>. This dataset consists of 11,807 instances (3,307 negative and 8,500 positive). The studies related to this dataset are referenced in [9] and [10].

Examples of the *cricket*, *restaurant*, and *drama* datasets are shown in Tables 1, 2, and 3, respectively.

Table 1. Examples of data and class labels in the *cricket* domain. For non-native Bangla readers, the corresponding English-translated sentences are provided for their understanding which is absent in the dataset.

SL. No.	Bangla Sentence	English Translation	Class (Tag)
1	হাতুরি একটা হিংসুটে লোক	Haturi is a violent man	Negative (0)
2	আজকে ম্যাচ তামিমের জন্য হারবো।	We will lose the match today for Tamim.	Negative (0)
3	সাব্বিরের শেষের ছক্সা টা অসাধারণ লেগেছে।	Sabbir's last six was amazing.	Positive (1)
4	ভালোবাসি দেশ এবং ক্রিকেটকে	Love country and cricket	Positive (1)
5	প্রথম দিকে উইকেট না পেলে ম্যাচ বাঁচানো কষ্টের হবে।	It will be difficult to save the match if we don't get early wickets.	Neutral (2)

Table 2. Examples of data and class labels in the *restaurant* domain. For non-native Bangla readers, the corresponding English-translated sentences are provided for their understanding which is absent in the dataset.

SL. No.	Bangla Sentence	English Translation	Class (Tag)
1	স্টাফ কিন্তু, আমাদের জন্য ভয়ঙ্কর ছিল।	The staff however, was horrible for us.	Negative (0)
2	এটি খুব বেশি চিত্তাকর্ষক এবং খুব সুস্বাদু নয়।	It is not very impressive and very tasty.	Negative (0)
3	এটা চমত্কার ও খুব সস্তা ছিল	It was fantastic and very cheap	Positive (1)
4	চমৎকার ওয়াইন তালিকা	Excellent wine list	Positive (1)
5	আমি সেখানে বসে ছিলাম	I was sitting there	Neutral (2)

Table 3. Examples of data and class labels in the *drama* domain. For non-native Bangla readers, the corresponding English-translated sentences are provided for their understanding which is absent in the dataset.

SL. No.	Bangla Sentence	English Translation	Class (Tag)
1	বাজে অভিনয়	Bad acting	Negative (0)
2	ভাল লাগলো না	It didn't feel good	Negative (0)
3	ফালতু একটা নাটক	A silly drama	Negative (0)
4	অনেক সুন্দর নাটক	Very nice drama	Positive (1)
5	চমৎকার কাহিনী।	Nice story.	Positive (1)

### 3. RELATED WORK

Previous investigations into sentiment analysis in the Bangla language have predominantly been based on traditional machine learning techniques coupled with feature extraction methods. For instance, Sharif et al. [11] utilized Multinomial NB (MNB) for sentiment analysis. Bhowmik et al. [12] harnessed a Support Vector Machine (SVM) integrated with an expanded lexicon dictionary. Alshari et al. [13] applied a sentiment lexicon dictionary based on word2vec for sentiment classification. Chowdhury and Chowdhury [14] opted for a semi-supervised bootstrapping technique utilizing SVM and maximum entropy classifiers, relying on SentiWordNet for translation. Hasan et al. [15] introduced an XML-based POS tagger and SentiWordNet for valency analysis. Islam et al. [16] derived sentiment polarity by tokenizing adjective words using a POS tagger and identifying valence-shifting negative words. Tuhin et al. [17] engineered an emotion detection system by associating texts with emotion classes. Tabassum and Khan [18] devised a framework centered on tallying positive and negative words. Akter and Aziz [19] designed a lexicon-based dictionary model. Subsequently, Bhowmik et al. [20] experimented with various deep-learning architectures and hybrid models on a Bangla dataset for sentiment analysis. Moreover, Al-Mahmud and Shimada [21] introduced several transformers-based combined approaches for Aspect-Opinion Extraction in the Bangla fine-grained sentiment analysis. Luo et al. [3] proposed a modified contrastive objective with in-batch negative samples so that the sentence representations from the same class will be pushed close while those from the different classes become further apart in the latent space. Du et al. [4] designed a post-training procedure that contains the target domain masked language model (MLM) task and a novel domain-distinguish pre-training task. The post-training procedure encouraged BERT to be domain-aware and distilled the domain-specific features in a self-supervised way. Based on this, adversarial training was conducted to derive the enhanced domain-invariant features. Wu and Shi [5] proposed a novel Adversarial Soft Prompt Tuning method (AdSPT). On the one hand, AdSPT adopted separate soft prompts instead of hard templates to learn different vectors for different domains, thus alleviating the domain discrepancy of the [MASK] token in the MLM task. On the other hand, AdSPT used a novel domain adversarial training strategy to learn domain-invariant representations between each source domain and the target domain.

Our research diverges from traditional methodologies with machine-learning-based sentiment analysis by adopting transformers-based models, aligning with the current trend in sentiment analysis research. Renowned for their adeptness in capturing intricate linguistic structures, transformers-based models, namely BERT typically surpass traditional machine-learning-based models. Leveraging transformers-based models, we propose a domain-invariant transfer learning approach that is effective in sentiment analysis tasks, especially in the Bangla cross-domain.

## 4. METHOD

In the following subsections, we will discuss the basic models, baselines, and proposed approaches.

### 4.1. Basic Model

#### 4.1.1. Machine-Learning-Based Model

We utilized support vector machine (SVM) and random forest (RF) classifiers as machine-learning-based models because they are well-known as stronger classifiers in the traditional approaches. Feature extraction is performed using the term frequency-inverse document frequency (TF-IDF) with *sklearn*, a Python-based machine-learning library.

##### 4.1.1.1. Support Vector Machine (SVM)

The study [22] introduced SVM as a powerful model for classification tasks. SVM constructed a hyperplane in a high-dimensional space to separate data points of different classes by maximizing the margin between the classes. This is achieved by solving a convex optimization problem with linear constraints. The study [22] proved strong theoretical guarantees and practical effectiveness of SVM for the classification task.

##### 4.1.1.2. Random Forest (RF)

The study [23] introduced RF as an ensemble learning model that constructs multiple decision trees during training and merges their predictions. Each tree is trained on a random subset of the data, and the final prediction is made by aggregating the outputs (i.e., by majority voting from the predicted output). This approach improved the performance by reducing overfitting and variance compared to individual decision trees.

#### 4.1.2. Transformers-Based Model

We utilized mBERT, BanglaBERT, and BanglishBERT models as transformers-based models. They are popular and effective for the Bangla language tasks with less parameter utilization as compared to other transformers-based models.

##### 4.1.2.1. mBERT

mBERT, an abbreviation for multilingual bidirectional encoder representations from transformers, is a language model pre-trained on an extensive corpus of text encompassing 104 different languages worldwide [24]. Similar to its predecessor, namely BERT, mBERT undergoes pre-training via masked language modeling (MLM) and next sentence prediction (NSP) tasks. This pre-training process endows mBERT with a robust comprehension of multilingual contexts and linguistic structures.

In practical applications, fine-tuning the pre-trained mBERT model involves the addition of just one output layer, typically in the form of linear layers, to adapt it to specific downstream tasks. In our scenario, we incorporate linear layers atop mBERT for our sentiment analysis task in the Bangla language. This fine-tuning procedure enables mBERT to effectively capture and analyze sentiment patterns within Bangla text, leveraging its comprehensive multilingual pre-training.

#### **4.1.2.2. Bangla BERT**

BanglaBERT is a BERT-based natural language understanding (NLU) model designed specifically for the Bangla language, aimed at overcoming the challenges posed by its low-resource nature [25]. Similar to BERT, BanglaBERT is versatile and can be fine-tuned to address various downstream tasks in the Bangla language. To customize BanglaBERT for our particular sentiment analysis task, we incorporate linear layers on top of the model for fine-tuning. This fine-tuning procedure empowers BanglaBERT to accurately comprehend and analyze sentiment tasks.

#### **4.1.2.3. BanglishBERT**

BanglishBERT is a bilingual model pre-trained on both Bangla and English languages simultaneously, using the same set of hyperparameters as specified in BanglaBERT [25]. This innovative model excels in training tasks within the English language while also possessing the ability to transfer its knowledge to tasks in the Bangla language. Like our utilization of mBERT and BanglaBERT, we leverage BanglishBERT for our sentiment analysis task.

### **4.2 Baseline**

#### **4.2.1. Without Source Data Approach**

We set without source data (i.e., only utilizing the target data) approach by employing the above-mentioned machine-learning-based and transformers-based models. Therefore, there is no effect of source data in the training in this scenario. The model learns and evaluates solely based on the target data like traditional methods. To this end, this approach utilizes relatively less data during training compared to other baselines and the proposed method.

#### **4.2.2. Combined Data Approach**

**Step I:** Split the target data into training and test sets.

**Step II:** Merge the training sets of the target data with the source data.

**Step III:** Train (and validate) the model with the merged data.

**Step IV:** Perform inference using the test set of the target data.

This approach is performed by utilizing both machine-learning-based and transformers-based models.

#### **4.2.3. Stepwise Learning Approach**

Employing a stepwise learning approach, we leverage both source and target data to boost model performance for the main task. This method capitalizes on auxiliary information from source tasks, particularly beneficial when data for the target tasks are limited. The process involves two stages: source and target tasks. In the first stage, we train a transformers-based model, such as BanglaBERT, using the source data. Subsequently, in the second stage, we refine the model acquired from the first stage through re-training and evaluate its efficacy using the target dataset. For a comprehensive understanding of the stepwise learning procedure, we refer to the study [26]. Notably, this strategy utilized transformers-based models but not the conventional machine-learning-based models [26].

### 4.3. Proposed Method: Domain-Invariant Transfer Learning

Among the baseline approaches, namely combined data and stepwise learning approaches may improve the result if the two datasets (source and target) are in the same domain with task compatibility and/or label sharing between the datasets/tasks. However, our study deals with two different domain datasets (i.e., cross-domain) although compatibility and/or label sharing exists between the datasets/tasks. Therefore, there is a domain gap between these two datasets, and need to mitigate the gap through learning the model in a better way. To this end, we need to devise a method that can able to bridge this gap.

Studies [27, 28] suggested that BERT's final encoder layers are highly specialized for specific tasks, whereas the initial layers mainly encode positional information, and the middle layers focus on dependency relations. Therefore, we conceived and utilized this information from the studies [27, 28] for our proposed domain-invariant transfer learning approach. The proposed approach consists of two stages. In the first stage, we fine-tuned only the final encoder layer of BERT with the merged source, and training and validation sets of the target datasets to prioritize task-specific knowledge, avoiding an emphasis on domain-specific or structural language elements. This technique bridges the domain gap by focusing on task-specific knowledge when integrating the source and target (train and validation sets) datasets. Now the model has task-specific domain-invariant knowledge from two different domains. The acquired knowledge or parameters are then adapted for the target task in the second stage using the target dataset. In this stage, the updated model is re-trained and evaluated utilizing the target dataset. Note that we did not use the test set of the target data in the first stage because of the test/unseen data leakage for the second stage for model evaluation.

In domain-invariant transfer learning, the target data is utilized twice: (1) in the first stage (only training and validation sets) with the source data to mitigate the domain gap (i.e., to learn the domain-invariant features of the source and target domains) and (2) in the second stage to re-train and evaluate the model.

The proposed approach differs from existing stepwise learning [26] where the target data is utilized once because the source and target data are in the same domain. Therefore, there was no need to fine-tune task-specific encoder layers by the target data (training and validation sets) merging with source data to mitigate the domain gap in the first stage. Note that we did not fine-tune all encoder layers of BERT in the first stage because fine-tuning all encoder layers may degrade performance as the dataset is relatively small for fine-tuning. Moreover, our focus is to reduce the domain gap by utilizing the task-specific encoder layer.

The domain-invariant transfer learning comprises the following six steps.

**Step I:** Add linear layers on the top encoder layer of mBERT/BanglaBERT/BanglishBERT. We only needed the model's output for the [CLS] token (i.e., classification token) tensor for fine-tuning/training as both source and target tasks are sentence-level (sentiment) classifications. Hence, we select that slice of the [CLS] tensor cube and discard other token tensors during fine-tuning/training.

**Step II:** Fine-tune/train and update the parameters only for the last encoder layer of mBERT/BanglaBERT/BanglishBERT with the merged source, and training and validation sets of the target datasets. The rest of the encoder layers are kept frozen.

**Step III:** Freeze the last encoder layer of mBERT/BanglaBERT/BanglishBERT.

**Step IV:** Re-use the fine-tuned/trained model and replace the linear layer on top of BERT with the same output label of the target data/task.

**Step V:** Fine-tune/train only the linear layer of the updated model with the training and validation

sets of the target data, but not the mBERT/BanglaBERT/BanglishBERT encoder layers (i.e., keep all the layers frozen).

**Step VI:** Evaluate the model with the test set of the target data.

This technique consists of two stages. Steps I-V belong to the first stage, and steps V and VI belong to the second stage. Note that similar to stepwise learning [26], the proposed method utilizes the BERT-based models, namely mBERT, BanglaBERT, and BanglishBERT, but not the machine-learning-based models.

The overview of the proposed domain-invariant transfer learning approach is illustrated in Figure 1.

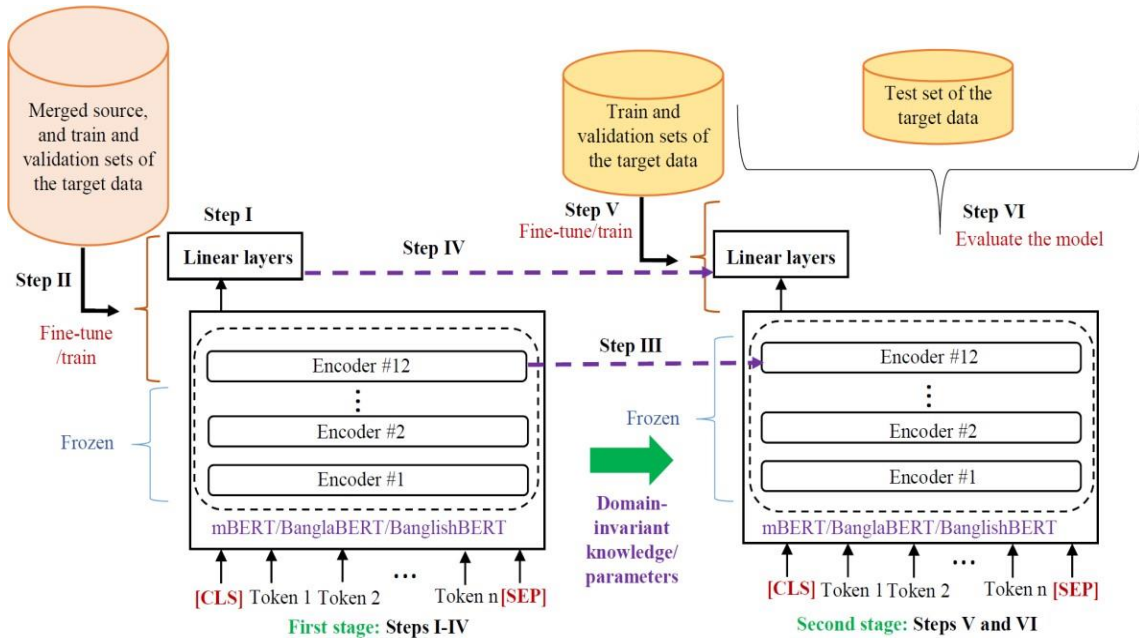


Figure 1. Overview of proposed domain-invariant transfer learning approach.

## 5. EXPERIMENT AND ANALYSIS

There are a total of six combinations of our source and target data: *drama & cricket*, *drama & restaurant*, *cricket & drama*, *restaurant & drama*, *restaurant & cricket*, *cricket & restaurant*. Here the left side of the symbol “&” is considered as the “source data” and the right side is regarded as the “target data”. To verify the reliability and robustness of the proposed method across all the domains, we looked at all six combinations in the experiment. If the proposed method performs better than the baselines in most combinations or all combinations, we can establish that the proposed method is effective across different domains or cross-domain datasets/tasks.

### 5.1. Baseline

We applied five baselines by utilizing without source data, combined data, and stepwise learning approaches through deploying machine-learning-based and transformers-based models. For without source data and combined data, we implemented support vector machine (SVM) and random forest (RF) for each of all the six combinations of the datasets. For without source data, combined data approach, and stepwise learning, we employed mBERT, BanglaBERT, and BanglishBERT for each of the six combinations of the datasets. Note that we only reported the



best score with the respective model as a summary among all the tested models for each of the six combinations of the datasets (shown in Table 4).

## 5.2. Proposed Approach

For the proposed domain-invariant transfer learning, we also utilized mBERT, BanglaBERT, and BanglishBERT. The proposed method solely utilized mBERT, BanglaBERT, and BanglishBERT, but not machine-learning-based models. Note that we only reported the best score with the respective model as a summary among the tested models for each of all the six combinations of the datasets (shown in Table 4).

## 5.3. Experimental and Hyper-Parameters Settings

The experiments were performed on a Linux server equipped with a CPU: Xeon E5-2620@2.10GHz 32 cores, Memory: 256GB, and GPU: Quadro RTX8000 (48GB), with the implementation carried out in Python 3.6. As the class distribution in each of the datasets is imbalanced, we used the weighted *F1* score as an evaluation metric.

The hyper-parameter settings are listed below:

- a. **Without source data approach by machine-learning-based models:**  
Data splitting = 90:10 and default settings.
- b. **Without source data approach by transformers-based models:**  
Data splitting = 80:10:10, epoch = 5, batch size = 32, learning rate = 1e-3, and optimizer = AdamW.
- c. **Combined data approach by machine-learning-based models:**  
Target data splitting = 90:10 (training:testing), the merged data (i.e., merging of source data with the training set of the target data) is allocated for training, and default settings.
- d. **Combined data approach by transformers-based models:**  
Target data splitting = 90:10 (training:testing), merged (i.e., merging of source data with the training set of the target data) data splitting = 80:20 (training:validation), epoch = 5, batch size = 32, learning rate = 1e-3, and optimizer = AdamW.
- e. **Stepwise learning:**  
For the source/auxiliary task, data splitting (training:validation) = 80:20, epoch = 3, batch size = 32, learning rate = 1e-3, and optimizer = AdamW.  
For the target task, data splitting = 80:10:10, epoch = 5, batch size = 32, learning rate = 1e-3, and optimizer = AdamW.
- f. **Domain-invariant transfer learning:**  
For Step I, merged data splitting = 80:20 (training:validation), epoch = 3, batch size = 32, learning rate = 1e-3, optimizer = AdamW.  
For Step IV, target data splitting = 80:10:10, epoch = 5, batch size = 32, learningrate = 1e-3, optimizer = AdamW.

## 5.4. Experimental Result Analysis

Table 4 presents the experimental results of the baselines and the proposed method for all six combinations of the datasets.

For without source data approach, the transformers-based model exhibited superior performance than the machine-learning-based model for all the combinations. Typically, the transformers-

based model is anticipated to outperform the machine-learning-based model due to their ability to handle long-range dependencies, capture contextual information, and pre-trained knowledge. The same thing happened in this experiment. However, this was not true in all cases for the combined data approach. The reason is the divergence of the domain (that we focused on the proposed approach). Thus, only a naïve combination is not enough to tackle such a situation although the combined data contained much data and in general, models perform better with much data.

Table 4. Experimental results of baselines and the proposed method.

<b>Approach</b>	<i>drama &amp; cricket (with the best model)</i>	<i>drama &amp; restaurant (with the best model)</i>	<i>cricket &amp; drama (with the best model)</i>	<i>restaurant &amp; drama (with the best model)</i>	<i>restaurant &amp; cricket (with the best model)</i>	<i>cricket &amp; restaurant (with the best model)</i>
Baseline: Without source data approach by machine-learning-based model	0.6833 (RF)	0.5567 (RF)	0.8867 (SVM)	0.8867 (SVM)	0.6833 (RF)	0.5567 (RF)
Baseline: Without source data approach by transformers-based model	0.6967 (BanglaBERT T and BanglishBERT T Jointly)	0.6133 (BanglaBERT)	0.8933 (BanglaBERT)	0.8933 (BanglaBERT)	0.6967 (BanglaBERT and BanglishBERT T Jointly)	0.6133 (BanglaBERT)
Baseline: Combined data approach by machine-learning-based model	0.6900 (SVM)	0.5633 (RF)	0.8800 (SVM)	0.8763 (SVM)	0.7392 (RF)	0.5621 (SVM)
Baseline: Combined data approach by transformers-based model	0.6300 (BanglaBERT T)	0.5333 (BanglaBERT)	0.9315 (BanglishBERT T)	0.9453 (BanglaBERT)	0.7311 (BanglishBERT T)	0.5394 (mBERT)
Baseline: Stepwise learning	0.7067 (mBERT)	0.5867 (BanglishBERT T)	0.9436 (BanglaBERT)	0.9410 (BanglishBERT T)	0.6994 (mBERT)	0.5465 (BanglishBERT T)
<b>Proposed: Domain-invariant transfer learning</b>	<b>0.7595</b> (BanglaBERT T)	<b>0.6478</b> (BanglishBERT T)	<b>0.9564</b> (BanglaBERT)	<b>0.9564</b> (BanglaBERT)	<b>0.7522</b> (BanglaBERT)	<b>0.6584</b> (BanglaBERT)

Stepwise learning only improved the performance in case of *drama & cricket* and *cricket & drama*, compared to without source data and combined data approaches. The scores were 0.7067 and 0.9436, respectively. It underperformed for the other four combinations, namely *drama & restaurant*, *restaurant & drama*, *restaurant & cricket*, and *cricket & restaurant* compared to without source data and/or combined data case. To this end, we were also unable to establish the effectiveness of stepwise learning for cross-domain sentiment analysis.

However, for all six combinations of the datasets, the proposed domain-invariant transfer learning outperformed all the baselines. The reason is the model learned from both source and target data

in a domain-invariant nature from much data in the first stage and was able to transfer this domain-invariant knowledge to target data in the second stage (illustrated in Section 4.3). In other words, the first stage was able to mitigate the domain gap by learning the domain-invariant features from much data.

In comparison between *drama & cricket* and *drama & restaurant*, *drama & cricket* exhibited improved results in both baselines and the proposed method. The reason is the target *restaurant* data contained a significant number of instances as neutral (i.e., 1,221 out of a total of 2,059 mentioned in Section 2). However, the target *cricket* data contained a limited number of instances as neutral (i.e., 261 out of a total of 2,979 mentioned in Section 2). On the other hand, the source *drama* data did not have any neutral instances because the dataset was essentially prepared for binary classification: positive or negative (i.e., 0 neutral instances). Therefore, the model was unable to learn well the neutral class in the target *restaurant* data. However, the model was affected less in this case in the target *cricket* data due to a significantly lower number of neutral instances than in the target *restaurant* data. In addition, in most instances, sentiment expression or use of vocabulary/phrases for the target entity bore similarity across both *drama* and *cricket* datasets. For instance, in the *drama* dataset, the target entity was typically the “actors”, while in the *cricket* dataset, it was the “cricketers”. Therefore, opinion words or vocabulary/phrases used for targeting persons (i.e., actors and cricketers) were similar. However, no or less such kind of similarity existed between *drama* and *restaurant* datasets.

The results in *cricket & drama* and *restaurant & drama* demonstrated more improved scores than the rest of the four combinations in the proposed approach due to the less task complexity in the target *drama* data. As mentioned earlier, the target *drama* was prepared for a binary classification (positive or negative) task whereas the *cricket* and *restaurant* data was constructed for a 3-class (positive, negative, or neutral) classification task.

In comparison between *restaurant & cricket* and *cricket & restaurant*, *restaurant & cricket* yielded a better score in the proposed approach although both the datasets share all three labels: positive, negative, and neutral. The reason is the target *cricket* data contained more data than the target *restaurant* data (2,979 and 2,059, respectively mentioned in Section 2). Therefore, in the second stage of the proposed method, the model was re-trained and evaluated with relatively more data for the target task.

## 6. CONCLUSIONS

In this study, we conducted cross-domain sentiment analysis across different domain data in the Bangla language. By leveraging both source and target data, we employed five baseline approaches and a proposed approach, namely domain-invariant transfer learning. We performed extensive experiments in six combinations of source and target domain datasets, namely *drama & cricket*, *drama & restaurant*, *cricket & drama*, *restaurant & drama*, *cricket & restaurant*, and *restaurant & cricket*. Our findings concluded that the naïve combination of source and target data did not always improve the performance, but rather degraded performance due to domain divergence between them, as machine learning requires proper training knowledge from the data for optimal learning outcomes. Therefore, there was a necessity to bridge the gap between the source and target domains. Doing so by the proposed approach enabled the model to learn more effectively from both datasets, yielded improved results, and outperformed all the baselines.

Cross-lingual sentiment analysis is one of the most potential future research where the task is performed by transferring knowledge from typically high-resource language data (e.g., English) to low-resource language data (e.g., Bangla). Note that in this study, we performed the task in cross-domain within the same language but not cross-lingual.

## ACKNOWLEDGEMENTS

This work was supported by JST SPRING, Grant Number JPMJSP2154.

## REFERENCES

- [1] Al-Mahmud, & Shimada, K. (2023). Dataset Construction and Opinion Holder Detection Using Pre-trained Models. *IJSKM*, 7(2), pp. 1-17.
- [2] Li, J. & Hovy, E. (2017) Reflections on sentiment/opinion analysis, *In: A practical guide to sentiment analysis*, pp. 41–59.
- [3] Luo, Y., Guo, F., Liu, Z., & Zhang, Y. (2022). Mere Contrastive Learning for Cross-Domain Sentiment Analysis. *In Proceedings of the 29th International Conference on Computational Linguistics*, pp. 7099-7111.
- [4] Du, C., Sun, H., Wang, J., Qi, Q., & Liao, J. (2020). Adversarial and Domain-Aware BERT for Cross-Domain Sentiment Analysis. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4019-4028.
- [5] Wu, H., & Shi, X. (2022). Adversarial Soft Prompt Tuning for Cross-Domain Sentiment Analysis. *In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2438-2447.
- [6] Sen, O., Fuad, M., Islam, M. N., Rabbi, J., Hasan, M. K., Fime, A. A., Fuad, M. T. H., Sikder, D., & Iftee, M. A. R. (2021). Bangla Natural Language Processing: A Comprehensive Review of Classical, Machine Learning, and Deep Learning Based Methods. *CoRR*, abs/2105.14875.
- [7] Karim, M. A., Kaykobad, M., & Murshed, M. (2013). Technical Challenges and Design Issues in Bangla Language Processing. *IGI Global*. ISBN: 9781466639706.
- [8] Rahman, M. A., & Kumar Dey, E. (2018). Datasets for Aspect-Based Sentiment Analysis in Bangla and Its Baseline Evaluation. *Data*, 3(2), 15.
- [9] Sazzed, S., & Jayarathna, S. (2019). A sentiment classification in Bengali and machine-translated English corpus. *In 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 107-114.
- [10] Sazzed, S. (2020). Cross-lingual sentiment classification in low-resource Bengali language. *In Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pp. 50-60.
- [11] Sharif, O., Hoque, M. M., & Hossain, E. (2019). Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naïve Bayes. *In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pp. 1-6.
- [12] Bhowmik, N. R., Arifuzzaman, M., Hossain Mondal, M. R., & Islam, M. S. (2021). Bangla text sentiment analysis using supervised machine learning with extended lexicon dictionary. *Natural Language Processing Research*, 1(3-4), pp.34–45.
- [13] Alshari, E. M., Azman, A., Doraisamy, S., Mustapha, N., & Alkeshr, M. (2018). Effective method for sentiment lexical dictionary enrichment based on word2vec for sentiment analysis. *In 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, pp. 1–5.
- [14] Chowdhury, S., & Chowdhury, W. (2014). Performing sentiment analysis in Bangla microblog posts. *In 2014 International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 1–6.
- [15] Hasan, K. M. A., Rahman, M., & Badiuzzaman. (2014). Sentiment detection from Bangla text using contextual valency analysis. *In 2014 17th International Conference on Computer and Information Technology (ICCIT)*, pp. 292–295.
- [16] Islam, M. S., Islam, M. A., Hossain, M. A., & Dey, J. J. (2016). Supervised approach of sentimentality extraction from Bengali Facebook status. *In 2016 19th International Conference on Computer and Information Technology (ICCIT)*, pp. 383–387.
- [17] Tuhin, R. A., Paul, B. K., Nawrine, F., Akter, M., & Das, A. K. (2019). An automated system of sentiment analysis from Bangla text using supervised learning techniques. *In 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, pp. 360–364.
- [18] Tabassum, N., & Khan, M. I. (2019). Design an empirical framework for sentiment analysis from Bangla text using machine learning. *In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–5.
- [19] Akter, S., & Aziz, M. T. (2016). Sentiment analysis on Facebook group using lexicon-based approach. *In 2016 3rd International Conference on Electrical Engineering and Information*

*Communication Technology (ICEEICT)*, pp. 1–4.

- [20] Bhowmik, N. R., Arifuzzaman, M., & Hossain Mondal, M. R. (2022). Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms. *Array*, 13, pp. 100123.
- [21] Al-Mahmud, & Shimada, K. (2024). Dataset Construction and Evaluation for Aspect-Opinion Extraction in Bangla Fine-Grained Sentiment Analysis. *In 4th International Conference on Data Science and Application (ICDSA 2023)*, pp. 437–449.
- [22] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), pp. 273–297.
- [23] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), pp. 5–32.
- [24] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
- [25] Bhattacharjee, A., Hasan, T., Ahmad, W., Mubasshir, K. S., Islam, M. S., Iqbal, A., Rahman, M. S., & Shahriyar, R. (2022). BanglaBERT: Language Model Pretraining and Benchmarks for Low- Resource Language Understanding Evaluation in Bangla. *In Findings of the Association for Computational Linguistics: NAACL 2022*, pp. 1318-1327.
- [26] Al-Mahmud, & Shimada, K. (2023). Demonstration of Effectiveness of Nativeness in Stepwise Learning by Performing Sentiment Analysis. *In 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*, pp. 1-6.
- [27] Vig, J., & Belinkov, Y. (2019). Analyzing the structure of attention in a transformer language model. *arXiv preprint arXiv:1906.04284*.
- [28] Hoover, B., Strobel, H., & Gehrmann, S. (2019). exbert: A visual analysis tool to explore learned representations in transformer models. *arXiv preprint arXiv:1910.05276*.

## AUTHORS

**Al-Mahmud** received his B.Sc. (Engg.) and M.Sc. (Engg.) degrees from Khulna University of Engineering & Technology in 2010 and 2015, respectively. He is currently a Ph.D. student at the Department of Artificial Intelligence, Kyushu Institute of Technology. He is also an Associate Professor at the Department of Computer Science and Engineering, Khulna University of Engineering & Technology. His current research interests include natural language processing, especially opinion mining and sentiment analysis.



**Kazutaka Shimada** received his B.S., M.S., and Ph.D. degrees from Oita University in 1997, 1999, and 2003, respectively. He is currently a Professor at the Department of Artificial Intelligence, Kyushu Institute of Technology. His current research interests include natural language processing, especially sentiment analysis, summarization, and conversation understanding.

