Viva_Palestine at StanceNakba Shared Task: Actor and Topic-Aware Stance Detection in Public Discourse

Wafaa S. El-Kassas[1], Enas A. Hakim Khalil[2], Enas M.F. El Houby[2]

Faculty of Computers and Information Technology, The National Egyptian E-Learning University (EELU), Giza, Egypt[1], Systems & Information Department, National Research Centre, Dokki, Giza 12311, Egypt[2]

wafaa.elkassas@gmail.com, enaskhalil@gmail.com, enas_mfahmy@yahoo.com

Abstract

Recent research has increasingly focused on user-generated content to clarify opinions expressed in social media discourse. The Actor and Topic-Aware Stance Detection in Public Discourse challenge encourages research on stance detection in polarized social media discourse on the Palestinian–Israeli conflict. The challenge comprises two subtasks: one for actor-level alignments and the other for cross-topic generalization patterns. The StanceNakba2026 task includes two subtasks: (A) Actor-Level Stance Detection in English and (B) Cross-Topic Stance Detection in Arabic. Our team participated in both subtasks with the name "Viva_Palestine". In Subtask A, the proposed method is based on the Bert-Base-Uncased model and achieved a Macro F1-score of 0.9190, placing 6th out of 13 teams. In Subtask B, the proposed method is based on the MARBERT model and achieved a Macro F1-score of 0.8724 (the top rank in the leaderboard), placing first out of 10 teams. These results show that the proposed modelling method performs well for both entity-specific stance alignment and strong cross-topic generalization.

Keywords: Actor-Level, cross-topic, Stance detection, Bert-Base-Uncased, MARBERT

1. Introduction

With social media widely used as a primary medium for communication and information sharing, analyzing user-generated content to understand public opinion has become a critical research area in Natural Language Processing (NLP). Stance, defined as the expression of a speaker's viewpoint and evaluation about a certain proposition (Biber and Finegan, 1988) is an important part of this method. Stance identification enables critical applications such as sentiment analysis, opinion mining, and social media monitoring, facilitating the examination of public discourse and societal trends, particularly in political and social contexts (AlDayel and Magdy, 2021) , (Geng et al., 2026). Recent work, including shared tasks and newly created datasets from social media, has expedited research in this domain by providing benchmark resources and evaluating diverse modelling approaches. These initiatives demonstrate the effectiveness of transformer-based models for multi-topic stance detection while highlighting ongoing challenges such as neutral stance classification and generalization to unseen topics (Schiller et al., 2021). In this regard, the StanceNakba 2026 Shared Task (Aldous et al., 2026) examines stance identification in highly divided social discourse about the Palestinian–Israeli conflict and related regional debates. The task is structured around two different goals. Subtask A (Actor-Level Stance Detection) looks at political alignment at the actor level by figuring out if writers have a Pro-Palestine, Pro-Israel, or Neutral stance on the issue as a whole. Subtask B (Cross-Topic Stance Detection) focuses on topic-dependent stance classification, assigning Favor, against, or neither labels to opinions expressed about specific conflict-related issues, such as diplomatic normalization with Israel and the presence of refugees in Jordan. Stance identification is important across languages. StanceNakba 2026 works in both English and Arabic. Subtask A is completed in English, while Subtask B is completed in Arabic.

The remainder of this paper is organized as follows. Section 2 reviews related work on stance detection. Section 3 describes the proposed methodologies for both subtasks. Section 4 presents the experimental results. Finally, Section 5 concludes the paper and outlines directions for future work.

2. Related Work

Stance detection research aims to detect the position expressed by authors toward specific targets in social media content and has been explored through text-based, multilingual, and multimodal approaches. Early work stressed linguistic and rule-based methods, such as (Reveilhac and Schneider, 2023) which proposed a rule-based stance detection model for the SemEval Task 6A English dataset covering five topics, using sentiment lexical and linguistic markers, and achieving an average accuracy of 75%.

More recent studies have used deep learning and multimodal representation (AlShenaifi and Alangari, 2025) , introducing the MAWQIF-MM corpus of tweet–image pairs labeled with Favor, Against, and Neutral for Arabic multimodal stance detection. Using an attention-based model that combines AraBERT text embeddings with BLIP visual features, the approach achieved 88% accuracy, outperforming text-only and image-only baselines.

Proceedings of the 2nd International Workshop on Nakba Narratives as Language Resources (Nakba-NLP 2026) @ LREC 2026, pages 239–243
11 May 2026.

Beyond direct stance classification, several works have examined public opinion and polarization in conflict-related discourse and analyzed Chinese-language Weibo posts on the October 2023 Israeli-Palestinian conflict. Findings show netizens' mixed emotional responses and support for humanitarian concerns, largely aligning with the Chinese government's diplomatic positions, highlighting Weibo's role in shaping and reflecting public opinion. Similarly, (Guerra et al., 2025) investigated over 450,000 Reddit posts on the 2023 Israeli–Palestinian conflict using an unsupervised lexicon-based extremism score. Results showed that extreme sentiment increased by up to 80% during major conflict events, highlighting how real-world developments drive polarized opinions on social media.

Other studies have applied stance detection in domain-specific contexts (Britt et al., 2025) analyzed 4,245 posts from 1,319 users on a major U.S. trucking forum and showed that drivers predominantly expressed negative stances toward autonomous trucks, driven by concerns over feasibility and job displacement. Their work demonstrated the value of integrating stance detection with topic modelling, sentiment analysis, and emotion analysis to capture nuanced public opinions.

In contrast to prior work, this paper addresses stance detection in two distinct settings: actor-level detection in English and cross-topic detection in Arabic. Unlike previous studies that focus on a single language or task, this work evaluates performance across different languages and task formulations, highlighting challenges in both high-resource and low-resource scenarios.

3. Methodology

This section describes the methodology adopted for StanceNakba 2026 Subtasks A and B. It begins with a description of the datasets used in both subtasks, followed by detailed explanations of the proposed approaches for Subtask A and Subtask B, respectively.

3.1 Data

The dataset for StanceNakba 2026 Subtask A (Actor-Level Stance Detection) comprises 1,401 English-language social media posts, divided into 980 training samples, 210 validation samples, and 211 test samples. Each instance includes three fields: a unique text identifier, the post content, and a stance label indicating whether the post expresses a Pro-Palestine, Pro-Israel, or Neutral stance. Examples of the 3 classes are shown in Figure 1.

In contrast, the dataset for Subtask B (Cross-Topic Stance Detection) comprises 1,205 Arabic social media posts, with 843 training samples, 181 validation samples, and 181 test samples. Each record consists of four attributes: an identifier, the sentence text, the associated topic, and a stance label. Figure 2 illustrates examples of different classes. The labels classify each Sentence as Favor, against, or neither with respect to one of two predefined topics: Normalization with Israel (577 samples) and Refugee/Immigrant Presence in Jordan (628 samples). The data for this subtask are taken from **MARASTA** corpus (Charfi et al., 2024).

Pro-Palestine
"The systematic displacement of Palestinian families from their ancestral homes represents a clear violation of international law and the right of return."

Pro-Israel
"Israel's defensive measures are necessary responses to existential threats, ensuring the safety of its citizens against terrorism."

Neutral
"The conflict involves competing territorial claims, with both populations having deep historical connections to the region."

Figure 1: Examples of different classes from the dataset of Subtask A.

Favor
التطبيع مع إسرائيل سيحسن الاقتصاد ويجلب الاستثمارات للمنطقة

Against
التطبيع مع إسرائيل يعتبر خيانة لقضية فلسطين والشعب الفلسطيني

Neither
التطبيع مع إسرائيل هو قضية معقدة لها جوانب اقتصادية وسياسية متعددة

Figure 2: Examples of different classes from the dataset of Subtask B.

3.2 Methodology of Subtask A (English)

This task requires identifying the general position of a text author regarding the Palestinian-Israeli conflict, categorized as Pro-Palestine, Pro-Israel, or Neutral. The Bert-Base-Uncased model is used for 3-class stance classification (Neutral, Pro-Israel, Pro-Palestine), Figure 1 illustrates examples of different classes. This model is **uncased**, meaning it does not differentiate between uppercase and lowercase letters. For example, it treats the words "english" and "English" as the same. BERT (Bidirectional Encoder Representations from Transformers) is pretrained on a large corpus of English data using a self-supervised approach. This means it was trained on raw texts without human labeling, using an automatic process to generate inputs and labels.

First, the data was tokenized using the tokenizer from the pretrained model with max_length = 256 and padding = max_length.

In the training phase, the model was trained on 980 samples and tested on 210 validation samples.

In the testing phase, the model was trained on 1190 samples (a combination of 980 training samples and 210 validation samples) and tested on the 211 unlabeled test samples.

3.3 Methodology of Subtask B (Arabic)

In Subtask B, it is required to build a unified model to predict Favor, against, or neither stance toward specific conflict-related topics. The topics covered in Subtask B are: (1) normalization with Israel and (2) refugee presence in Jordan. Examples of the three classes are illustrated in Figure 2.

In the training phase, the model "aubmindlab/bert-base-arabertv02" was utilized. The model was trained on a dataset of 843 samples and evaluated on a validation set consisting of 181 samples.

In the testing phase, three models have been used with extensive experimentation conducted to fine-tune their hyperparameters. The models include "CAMeL-Lab/bert-base-arabic-camelbert-mix", "UBC-NLP/MARBERTv2", and "UBC-NLP/MARBERT". These models were trained on a total of 1024 samples (i.e., a combination of 843 training samples and 181 validation samples) and were subsequently evaluated on the 181 unlabeled test samples.

At the end of the test phase experiments, the MARBERT model "UBC-NLP/MARBERT" achieves the best results over the other models and secures the top rank (i.e., first place in the leaderboard) in Subtask B among 10 teams.

MARBERT (Abdul-Mageed et al., 2021) is a powerful Arabic-specific Transformer-based language model. It was trained on large-to-massive datasets that cover different text genres (including social media) and different domains.

4. Experimental Results

4.1 Results of Subtask A (English)

In the training phase, the bert-base-uncased model was trained using 980 examples and tested with the validation examples of 210 samples. Table 1 shows the hyperparmeters tuned and performance metrics obtained. In the testing phase, the model was trained using a combination of train and validation samples. Table 2 shows the hyperparmeters tuned and performance metrics obtained. The confusion matrix of testing model is shown in Figure 3. The confusion matrix reveals that most misclassifications occur between the Neutral class and the two stance-bearing classes. 6 Neutral instances are misclassified as Pro-Israel, and 4 Neutral instances are misclassified as Pro-Palestine.

Train data	980 samples
Test(validation)	210 samples
Learning rate	3e-05
epochs	20
Early stopping	Yes, stop at 17th
Accuracy	0.890
precision	0.897
recall	0.890
Macro F1	0.891

Table 1: The training phase results with the parameters tuned for Subtask A.

Train data	1190 (combination of train and validation)
Test	211 unlabeled test samples
Learning rate	10e-05
epochs	20
Early stopping	Yes, stop at 9th
Test Accuracy	0.9149
precision	0.9211
Recall	0.9193
Macro F1	0.9190

Table 2: Test phase results with parameters tuned for Subtask A.

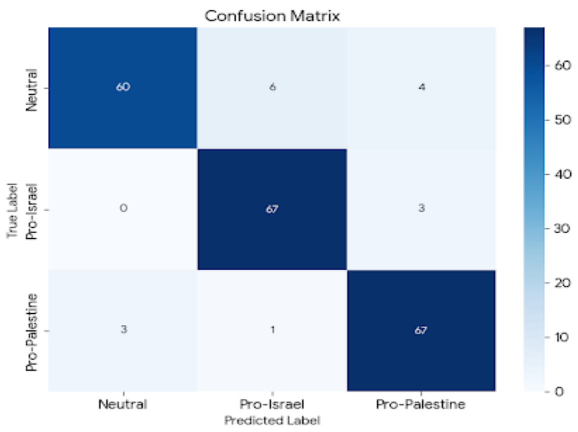

Figure 3: Confusion matrix for Subtask A.

The model is highly effective in detecting explicit stance expressions, achieving high recall for both Pro-Israel and Pro-Palestine classes. However, the comparatively lower recall for the Neutral class (85.7%), as illustrated in Table 3, highlights a limitation in capturing true neutrality. Overall, the model demonstrates robust and balanced performance, with particularly strong results in distinguishing between opposing stance categories. The primary challenge remains improving the detection of neutral content in the presence of subtle or implicit opinions.

class	Precision	recall	F1	support
Neutral	0.95	0.86	0.90	70
Pro-Israel	0.91	0.96	0.93	70
Pro-Palestine	0.91	0.94	0.92	71
Accuracy			0.92	211
Macro Avg	0.92	0.92	0.92	211
Weighted Avg	0.92	0.92	0.92	211

Table3: Classification Report for Subtask A.

4.2 Results of Subtask B (Arabic)

In the testing phase, the training data increased by merging the training samples with the validation samples to form a new training set that includes 1024 records. Three Transformer-based models for Arabic have been used in the testing phase.

Table 4 illustrates the tuning of hyperparameters for the models: “UBC-NLP/MARBERT”, “CAMeL-Lab/bert-base-arabic-camelbert-mix”, and “UBC-NLP/MARBERTv2”. Table 5 shows the results of these models based on the tuning of hyperparameters in Table 4. Many trials to tune the hyperparameters of these models have been done, but not all the results of these trials are included in this paper due to space constraints.

The MARBERT model "UBC-NLP/MARBERT" achieved the best results over the other two models, as shown in Table 5. The model demonstrates robust and well-balanced performance across all evaluation metrics. It is worth mentioning that the MARBERT model achieved first place (top rank) in the leaderboard for Subtask B among 10 teams.

5. Conclusion

In this paper, Viva_Palestine demonstrated how to identify an actor's stance in English and how to determine a stance across topics in Arabic. The experimental results show strong, reliable performance on both subtasks. Viva_Palestine placed sixth out of 13 teams and achieved a Macro F1-score of 0.9190 in the English actor-level track (Subtask A). More importantly, the system's Macro F1-score of 0.8724 placed it first in the Arabic cross-topic track (Subtask B). This shows that it can generalize across topics in a morphologically complex language.

The results indicate how well the suggested methods handle multilingual and cross-topic attitude identification in difficult real-world scenarios.

The limited dataset size in this study constrains the overall performance. Future improvements could be achieved by adopting alternative strategies such as applying cross-validation during training, leveraging data augmentation techniques, incorporating additional datasets, and fine-tuning a wider range of models.

MODEL_NAME	“UBC-NLP/MARBERT”, “CAMeL-Lab/bert-base-arabic-camelbert-mix”, and “UBC-NLP/MARBERTv2”
Training Data	1024 samples (combination of training set and validation set)
MAX_LENGTH	128
BATCH_SIZE	8
LEARNING_RATE	2e-5
NUM_EPOCHS	4
WARMUP_RATIO	0.1
WEIGHT_DECAY	0.01
SEED	42

Table 4: Settings of the different models for Subtask B.

MODEL_NAME	“CAMeL-Lab/bert-base-arabic-camelbert-mix”
accuracy	0.8011
precision	0.8025
recall	0.8016
macro_f1	0.8002
MODEL_NAME	“UBC-NLP/MARBERTv2”
accuracy	0.8729
precision	0.8710
recall	0.8712
macro_f1	0.8710
MODEL_NAME	“UBC-NLP/MARBERT”
accuracy	**0.8729**
precision	**0.8736**
recall	**0.8733**
macro_f1	**0.8724**

Table 5: Results of different models in the testing phase for Subtask B.

6. References

Abdul-Mageed, M., et al. (2021). ARBERT & MARBERT: Deep bidirectional transformers for Arabic. Proceedings of the 59th Annual Meeting of the ACL-IJCNLP 7088-7105.

AlDayel, A. and Magdy, W. (2021). Stance detection on social media: State of the art and trends. Information Processing & Management 58(4): 102597.

Aldous, K., et al. (2026). StanceNakba Shared Task: Actor and topic-aware stance detection in public discourse. Proceedings of the 15th International Conference on Language Resources and Evaluation (LREC'26).

AlShenaifi, N. and Alangari, N. (2025). Beyond text: Multimodal stance detection in Arabic tweets. Machine Learning with Applications 100823.

Biber, D. and Finegan, E. (1988). Adverbial stance types in English. Discourse Processes 11(1-34).

Britt, N., et al. (2025). Truck drivers and autonomous trucks: a topic modeling analysis of truck driver posts. Transportation Research Interdisciplinary Perspectives 34.

Charfi, A., et al. (2024). MARASTA: A multi-dialectal Arabic cross-domain stance corpus. *Joint International Conference on Computational Linguistics, Language Resources and Evaluation* 11060-11069.

Geng, G., et al. (2026). PERStance: Personality-guided enhanced multimodal stance detection. Information Processing & Management 63(4): 104593.

Guerra, A., et al. (2025). Quantifying extreme opinions on Reddit amidst the 2023 Israeli–Palestinian conflict. Natural Language Processing Journal 11(100156).

Reveilhac, M. and Schneider, G. (2023). Replicable semi-supervised approaches to state-of-the-art stance detection of tweets. Information Processing & Management 60(2): 103199.

Schiller, B., et al. (2021). Stance detection benchmark: How robust is your stance detection? KI – Künstliche Intelligenz 35(3): 329-341.