Quasar at MAHED Shared Task : Decoding Emotions and Offense in Arabic Text using LLM and Transformer-Based Approaches

Md Sagor Chowdhury, Adiba Fairooz Chowdhury
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology, Bangladesh
{u2004010, u2004014}@student.cuet.ac.bd

Abstract

The escalating presence of propaganda and hate speech on social media platforms underscores the need for robust automated detection systems to preserve the integrity of public discourse. Our team, participated in the MAHED 2025 Shared Task at the ArabicNLP 2025 conference, co-located with EMNLP 2025, focusing on Subtask 1 (Text-based Hate and Hope Speech Classification) and Subtask 2 (Emotion, Offensive, and Directed Hate Detection) in Arabic content. In Subtask 1, we experimented with models including XLM-RoBERTa-Large, Davlan/xlm-roberta-base-finetuned-arabic, asafaya/bert-base-arabic, aubmindlab/bert-base-arabertv2, Google Gemma-7B, and Qwen2.5-14B-Instruct, achieving the highest macro-f1 of 0.674 with Gemma-7B and ranking 12th on the leaderboard. In Subtask 2, using models such as aubmindlab/bert-base-arabertv2, Google Gemma-7B, Qwen2.5-14B-Instruct, asafaya/bert-base-arabic, and domain-specific hate-speech models, our best macro-f1 was 0.48 with both Gemma-7B and aubmindlab/bert-base-arabertv2, placing us 6th in the leaderboard.

1 Introduction

Hate and hope speech uses negative or positive expressions in text to influence readers' behavior, opinions, or emotions for a specific agenda. It is widespread on social media in tweets, posts, and comments, often with inherent bias. These speeches shape public perception and attract attention by amplifying offensive content, emotional appeals, or harmful narratives. Detecting hate, hope, and offensive content is essential to curb misleading or harmful information. Hate, hope, and offensive speech detection in Arabic text is challenging due to subtle sentiment, sarcasm, and context-dependent meanings. Social media content includes slang, abbreviations, and mixed styles, adding complexity. There is a gap in large-scale annotated datasets and specialized NLP tools for this compared to general sentiment analysis. This paper aims to detect such speech in Arabic social media posts and comments.

The MAHED 2025 shared task (Zaghouani et al., 2025) provides datasets for Subtask 1 and Subtask 2, labeled for offensive, hate, and hope speech, as a benchmark.We participated in subtask 1 and subtask 2. To achieve our goal, we augmented under-represented classes using back translation and evaluated models like XLM-RoBERTa-Large (Conneau et al., 2020), Davlan/xlm-roberta-base-finetuned-arabic (Davlan Team, 2023), asafaya/bert-base-arabic (Safaya et al., 2020), aubmindlab/bert-base-arabertv2 (Antoun et al., 2020), Google Gemma-7B (Mesnard et al., 2024) with classification head, and Qwen2.5-14B-Instruct (Yang et al., 2024). Each model was trained and assessed on the dataset. For Subtask 1 (Text-based Hate and Hope Speech Classification), Google Gemma-7B achieved a macro-F1 score of 0.674, ranking 12th. For Subtask 2 (Emotion, Offensive, and Directed Hate Detection), Gemma-7B and aubmindlab/bert-base-arabertv2 scored 0.48 macro-F1, placing 6th.

The core contributions of our research work include:

1. augmenting underrepresented classes using back translation,

2. leveraging external datasets to enrich training data, and

3. applying both large language models (LLMs) and Arabic-specific transformer models to improve detection of hate, hope, offensive, emotion, and directed hate speech in Arabic content.

Proceedings of The Third Arabic Natural Language Processing Conference, pages 692–699
November 8-9, 2025

Detailed implementation information is available in the linked GitHub repository [1]

2 Related Work

Hate speech detection in Arabic text has gained attention due to content moderation needs. Zaghouani et al. (2024a) developed a multi-label hate speech annotated Arabic dataset for model training. Biswas and Zaghouani (2025a) created an annotated corpus of Arabic tweets for hate speech analysis. This establishes benchmarks for Twitter-based systems. Hope speech serves as a counter to hate speech in research. Biswas and Zaghouani (2025b) introduced the EmoHopeSpeech dataset annotated for emotions and hope speech in English and Arabic. The bilingual dataset enables cross-lingual studies. It supports identifying positive and harmful content, offering a nuanced approach beyond binary classification. Hate speech research now includes multimodal analysis. Alam et al. (2024a) analyzed Arabic memes for propaganda-hate links using multi-agent LLMs. The ArMeme dataset (Alam et al., 2024b) shows how propagandistic memes evolve into hateful content. This highlights progression from persuasion to explicit hate. Propaganda detection intersects with hate speech. The WANLP 2022 shared task (Alam et al., 2022) set benchmarks for Arabic propaganda. Hasanain et al. (2024b) examined GPT-4 for propaganda spans in news. SemEval-2024 Task 4 (Dimitrov et al., 2024) focused on multilingual persuasion in memes. ArAIEval (Hasanain et al., 2023) targeted persuasion and disinformation in Arabic. Transformer models improve Arabic classification. Models like XLM-RoBERTa and AraBERT are standard. LLMs such as Gemma and Qwen perform well in tasks. Hasanain et al. (2024a) showed LLM effectiveness in propaganda annotation. Multitask learning detects emotions, offensive language, and hate using shared representations. Arabic hate speech detection continues to face challenges including dialectal variations, cultural context sensitivity, and evolving online hate speech patterns. The MAHED 2025 shared task builds upon these foundations while addressing contemporary challenges in Arabic social media content moderation through combining traditional classification with modern large language models.

[1] GitHub Repository

3 Data

For Subtask 1, we used the dataset from the MAHED 2025 shared task (Zaghouani et al., 2025), consisting of Arabic social media posts labeled as *not_applicable*, *hope*, or *hate*, introduced in (Zaghouani et al., 2024b). The dataset is divided into training, development, and test sets, though specific split sizes are not detailed here. The training set includes 6,890 samples with notable imbalances: 3,697 *not_applicable*, 1,892 *hope*, and 1,301 *hate*, as shown in Figure 1.

For Subtask 2, we used the EmoHopeSpeech dataset (Zaghouani and Biswas, 2025), which contains 5,960 rows annotated with emotions, offensive and hate speech in English and Arabic.The distribution of frequecies for each label are given in Table A.

Examples of each label are given in section B

4 System

We have participated in subtask 1 and 2, which are an unimodal and multilabel text classification task. Figure 2 presents our proposed multi-output classification architecture for Arabic text analysis.

4.1 Data Augmentation

We have done data augmentation only for subtask-1.The original training dataset exhibited significant class imbalance with 3,697 not_applicable, 1,892 hope, and 1,301 hate instances. We implemented a multi-stage augmentation strategy to address this imbalance.

External Dataset Integration We incorporated additional datasets: 130 instances from an Arabic optimism dataset[2] for hope speech and additional hate speech samples from an external corpus[3].

Synonym Replacement and Back-Translation For the "hope" class, we applied Arabic synonym replacement using a comprehensive dictionary[4], preserving URLs, emojis, and punctuation while replacing content words. We generated 1,675 additional samples through this process. We further implemented back-translation (Arabic English Arabic) using Google Translate API: (1)

[2] https://data.mendeley.com/datasets/mcnzzpgrdj/2

[3] https://www.sciencedirect.com/science/article/pii/S2352340923008144

[4] https://github.com/mdanok/ArabicLT/master/csv/synonyms.csv

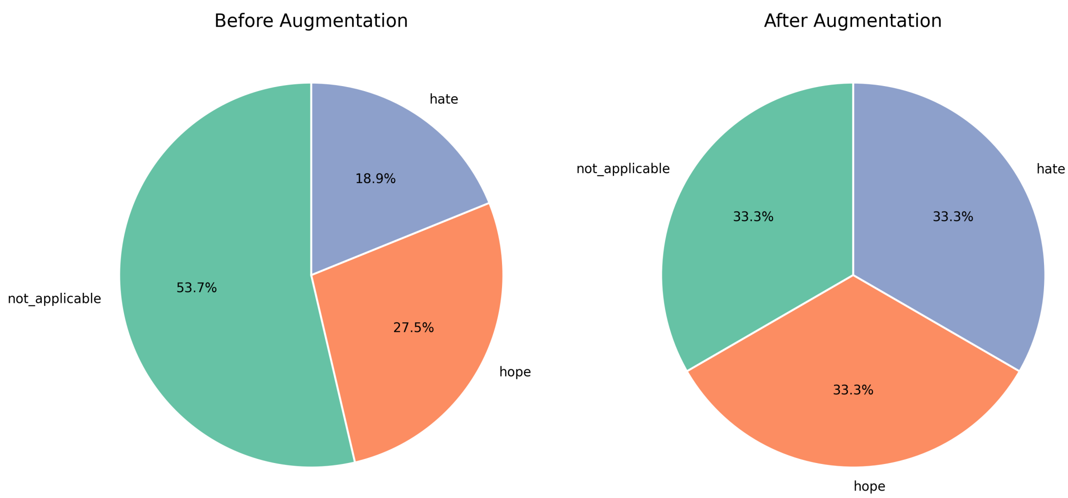

Figure 1: Label Distribution Before and After Augmentation for subtask-1

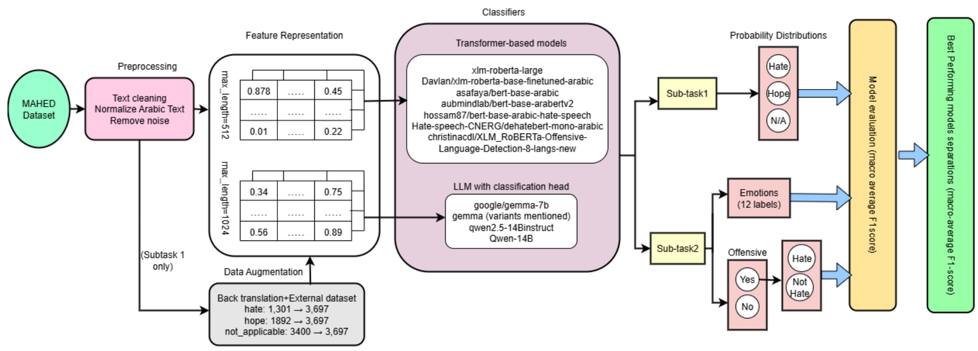

Figure 2: Overview of our proposed multi-output classification system for Arabic text analysis.

translating synonym-replaced Arabic to English, (2) applying English synonym replacement using WordNet and spaCy on nouns, verbs, adjectives, and adverbs, and (3) translating back to Arabic. This introduced natural linguistic variations while preserving semantic content.

Label	Before	After
not_applicable	3,697	3,697
hope	1,892	3,697
hate	1,301	3,697
Total	**6,890**	**11,091**

Table 1: Dataset distribution before and after augmentation

The augmentation successfully created a balanced dataset with 3,697 instances per class, representing a 61% increase in total samples and ensuring equal learning opportunities for all categories. Figure 1 illustrates the class distribution before and after the augmentation process.

4.2 Data Preprocessing

To ensure clean and consistent input for our models, we implemented a comprehensive preprocessing pipeline shown in this table2 for the Arabic text data. The preprocessing steps were applied sequentially to handle the specific challenges of Arabic social media text. This preprocessing pipeline ensured that our models received clean, normalized Arabic text optimized for classification tasks while preserving essential semantic content.

4.3 Initial Experimentation

In our initial experiments for both subtasks, we explored several transformer-based models to establish baselines and understand the performance landscape. Using `xlm-roberta-large` on both augmented+preprocessed and raw datasets, we observed that while the augmented data showed a strong bias towards the *hope* label, performance on the raw dataset was primarily skewed towards *not_applicable*. Balanced experiments on subsets demonstrated that data preprocessing

Before	Actions	After
RT @user123: أنا سعيد جداً 😊 *(RT @user123: We are very happy 😊)* http://example.com	Remove RT and mentions	نحن سعداء جداً 😊 *(We are very happy 😊)* http://example.com
أنا سعيد جداً 😊 *(We are very happy 😊)* http://example.com	Remove URLs	نحن سعداء جداً 😊 *(We are very happy 😊)*
نحن سعداء جداً 😊 *(We are very happy 😊)*	Emoji to Arabic translation	نحن سعداء جداً وجه مبتسم *(We are very happy smiling face)*
نحن سعداء جداً وجه مبتسم *(We are very happy smiling face)*	Remove remaining emojis	نحن سعداء جداً وجه مبتسم *(We are very happy smiling face)*
نَحْنُ سُعَدَاءُ جِدّاً وَجْهٌ مُبْتَسِمٌ *(We are very happy smiling face - with diacritics)*	Remove diacritics	نحن سعداء جداً وجه مبتسم *(We are very happy smiling face)*
نحن سعداء جداً وجه مبتسم!!! *(We are very happy smiling face!!!)*	Character filtering	نحن سعداء جداً وجه مبتسم *(We are very happy smiling face)*
نحن سعداء جداً وجه مبتسم في هذا *(We are very happy smiling face in this)*	Stopword removal	نحن سعداء جداً وجه مبتسم *(We are very happy smiling face)*

Table 2: Examples of Preprocessing Actions on Arabic Text.

and balancing played a crucial role in improving model performance. Building on this, we evaluated additional models including *Gemma*, *Qwen* (14B and 2.5-14B-instruct), *asafaya/bert-base-arabic*, *aubmindlab/bert-base-arabertv2*, and *Davlan/xlm-roberta-base-finetuned-arabic*, as well as specialized hate speech models such as *hossam87/bert-base-arabic-hate-speech* (Hossam, 2023) and *Hate-speech-CNERG/dehatebert-mono-arabic* (Aluru et al., 2020), which showed strong bias towards the *hate* label in training. Across these experiments, performances varied depending on model architecture and data preprocessing strategies.

4.4 Overview of the Adopted Model

For Subtask 1, we evaluated several models on different versions of the dataset, including XLM-RoBERTa-large, Qwen-14B, and Davlans XLM-RoBERTa-base fine-tuned models. Among these, Gemma7b with selected parameters(C) consistently achieved the highest accuracy across combinations of training, validation, and test sets, outperforming others with accuracies ranging from 0.47 to 0.67 depending on the dataset composition.

Similarly, for Subtask 2, Gemma was again chosen as the primary model. Other transformer-based models showed competitive performance, but Gemma provided the most consistent and reliable results for our multi-label Arabic classification task. We used the standard pre-trained tokenizer, set appropriate maximum sequence lengths, and experimented with hyperparameters such as learning rate, batch size, and number of epochs to optimize performance.

5 Results and Analysis

In this section, we summarize the key findings of our experiments, focusing on which approaches performed best rather than presenting full tables.

5.1 Evaluation Metrics

We evaluate using the macro-F1 score, which balances performance across all classes by averaging F1-scores independently for each label.

5.2 Comparative Analysis

Our experiments show that among all tested approaches, **Google Gemma-7B with DoRA configuration** achieved the best results for Subtask 1 (Hate and Hope Speech Classification), reaching an accuracy of **0.67**. The comprehensive performance comparison across different models and dataset configurations for Subtask 1 is presented in Table 3.

Model	Dataset Configuration	Macro-F1
XLM-RoBERTa-Large	Augmented + Preprocessed	0.33
	Given Dataset	0.54
	Given Dataset (1301 per label)	0.32
	Cleaned + 1301 per label + Non-cleaned Val	0.59
	Preprocessed + 1301 per label + Cleaned test	0.57
	Preprocessed given + Unprocessed test	0.23
Google Gemma-7B	Given + LoRA config	0.66
	Given + preprocessed test	0.60
	Given + 1301 per label	0.48
	Given + 1301 per label + processed test	0.47
	Given + DoRA + Unprocessed test	**0.67**
	Given + DoRA + processed test	0.64
Qwen-14B	1300 data samples	0.43
Davlan/xlm-roberta-base-arabic	Given Dataset	0.63
	Preprocessed Dataset	0.61
	Augmented + preprocessed	0.59

Table 3: Performance comparison for Subtask 1 across different models and dataset configurations.

For Subtask 2 (Emotion, Offensive, and Directed Hate Detection), the highest macro-F1 score (**0.48**) was obtained by three models: **Qwen2.5-14B-Instruct**, **aubmindlab/bert-base-arabertv2**, and **Google Gemma-7B**. The performance comparison for Subtask 2 is shown in Table 4.

6 Error Analysis

6.1 Confusion Matrix Analysis

To evaluate the classification performance in detail, we analyze the confusion matrices generated by our best performing Gemma-7B model. For Subtask 1, the confusion matrix (shown in Figure 8 in Appendix D) demonstrates the model's ability to distinguish between hate speech, hope speech, and not applicable content.

Model	Macro-F1	Notes
Qwen2.5-14B-Instruct	**0.48**	-
asafaya/bert-base-arabic (3 epochs)	0.45	-
asafaya/bert-base-arabic (20 epochs)	0.44	-
aubmindlab/bert-base-arabertv2	**0.48**	-
aubmindlab/bert-base-arabertv2	0.42	Preprocessed
Google Gemma-7B	**0.48**	-
Ensemble (XLM-RoBERTa + Gemma + dehatebert)	0.43	-

Table 4: Performance comparison for Subtask 2 showing macro-F1 scores.

For Subtask 2, we examine three separate confusion matrices corresponding to the multi-label classification components: emotion classification (Figure 9), offensive content detection (Figure 10), and hate speech detection within offensive content (Figure 11). These matrices provide insights into the model's performance across different aspects of the multi-label task.

6.2 Error Patterns

The confusion matrices reveal several key patterns in model performance:

- **Subtask 1 Performance**: The model shows good discrimination between hate and hope classes but occasionally confuses both with *not_applicable* content. This suggests that the model sometimes struggles to identify the presence of clear hate or hope indicators in ambiguous text.

- **Emotion Classification**: The model performs well on distinct emotions like joy and anger but struggles with subtle emotional distinctions. This indicates that while the model can capture clear emotional signals, it faces challenges in differentiating between closely related emotional states.

- **Offensive Content Detection**: The analysis shows high precision but some recall issues, particularly with borderline cases. This suggests the model tends to be conservative in its offensive content predictions, potentially missing some subtle forms of offensive language.

- **Hate Speech Detection**: Within offensive content, the model demonstrates the inherent challenge of distinguishing targeted hate from general offensive language. This highlights the complexity of the hate speech detection task, where the boundary between offensive and hateful content is often nuanced.

These error patterns provide valuable insights into the limitations of current approaches and suggest directions for future improvements in hate and hope speech classification systems.

7 Conclusion

Our study demonstrates the effectiveness of transformer-based and LLM approaches for Arabic hate, hope, offensive, and emotion detection, with Gemma-7B achieving the strongest results. However, the models show limitations in handling text with divine or religiously inspired speech, often misclassifying such hopeful expressions as neutral or humorous, as observed in our error analysis. Moreover, due to limited computational resources, we could not experiment with larger models capable of capturing broader context. As future work, we plan to incorporate domain-specific religious and cultural corpora to better model divine hopeful speech and explore larger-scale or more efficient models to enhance contextual understanding and overall robustness.

References

Firoj Alam, Md Rafiul Biswas, Uzair Shah, Wajdi Zaghouani, and Georgios Mikros. 2024a. Propaganda to hate: A multimodal analysis of arabic memes with multi-agent llms. In *International Conference on Web Information Systems Engineering*, pages 380–390. Springer.

Firoj Alam, Abul Hasnat, Fatema Ahmad, Md. Arid Hasan, and Maram Hasanain. 2024b. ArMeme: Propagandistic content in Arabic memes. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 21071–21090, Miami, Florida, USA. Association for Computational Linguistics.

Firoj Alam, Hamdy Mubarak, Wajdi Zaghouani, Giovanni Da San Martino, and Preslav Nakov. 2022. Overview of the WANLP 2022 shared task on propaganda detection in Arabic. In *Proceedings of the The Seventh WANLP*, Abu Dhabi, United Arab Emirates (Hybrid). ACL.

Sai Saket Aluru, Binny Mathew, Punyajoy Saha, and Animesh Mukherjee. 2020. Deep learning models for multilingual hate speech detection. *arXiv preprint arXiv:2004.06465*.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. AraBERT: A pre-trained arabic language model. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools*, Marseille, France. European Language Resource Association.

Md. Rafiul Biswas and Wajdi Zaghouani. 2025a. An annotated corpus of arabic tweets for hate speech analysis. *CoRR*, abs/2505.11969.

Md. Rafiul Biswas and Wajdi Zaghouani. 2025b. Emohopespeech: An annotated dataset of emotions and hope speech in english and arabic. *CoRR*, abs/2505.11959.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy. Association for Computational Linguistics.

Davlan Team. 2023. Xlm-roberta base fine-tuned for arabic. No specific academic paper is associated with Davlan/xlm-roberta-base-finetuned-arabic; refer to the Hugging Face model page for details.

Dimitar Dimitrov, Firoj Alam, Maram Hasanain, Abul Hasnat, Fabrizio Silvestri, Preslav Nakov, and Giovanni Da San Martino. 2024. Semeval-2024 task 4: Multilingual detection of persuasion techniques in memes. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, Mexico City, Mexico. Association for Computational Linguistics.

Maram Hasanain, Fatema Ahmad, and Firoj Alam. 2024a. Large language models for propaganda span annotation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 14522–14532, Miami, Florida, USA. Association for Computational Linguistics.

Maram Hasanain, Fatema Ahmed, and Firoj Alam. 2024b. Can gpt-4 identify propaganda? annotation and detection of propaganda spans in news articles. In *Proceedings of the 2024 Joint International Conference On Computational Linguistics, Language Resources And Evaluation*, LREC-COLING 2024, Torino, Italy.

Maram Hasanain, Firoj Alam, Hamdy Mubarak, Samir Abdaljalil, Wajdi Zaghouani, Preslav Nakov, Giovanni Da San Martino, and Abed Alhakim Freihat. 2023. ArAIEval Shared Task: Persuasion techniques and disinformation detection in arabic text. In *Proceedings of the First Arabic Natural Language Processing Conference (ArabicNLP 2023)*, Singapore. Association for Computational Linguistics.

Hossam. 2023. Bert base arabic hate speech detection model. Hugging Face model repository.

Thomas Mesnard, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, et al. 2024. Gemma: Open models based on gemini research and technology. *arXiv preprint arXiv:2403.08295*.

Ali Safaya, Moutasem Abdullatif, and Deniz Yuret. 2020. KUISAIL at SemEval-2020 task 12: BERT-based multi-label classification for offensive language detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, Barcelona (online). International Committee for Computational Linguistics.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, et al. 2024. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*.

Wajdi Zaghouani and Md Rafiul Biswas. 2025. Emohopespeech: An annotated dataset of emotions and hope speech in english and arabic. *arXiv preprint arXiv:2505.11959*.

Wajdi Zaghouani, Md Rafiul Biswas, Mabrouka Bessghaier, Shimaa Ibrahim, Georgios Mikros, Abul Hasnat, and Firoj Alam. 2025. MAHED shared task: Multimodal detection of hope and hate emotions in arabic content. In *Proceedings of the Third Arabic Natural Language Processing Conference (ArabicNLP 2025)*, Suzhou, China. Association for Computational Linguistics.

Wajdi Zaghouani, Hamdy Mubarak, and Md. Rafiul Biswas. 2024a. So hateful! building a multi-label hate speech annotated Arabic dataset. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15044–15055, Torino, Italia. ELRA and ICCL.

Wajdi Zaghouani, Hamdy Mubarak, and Md Rafiul Biswas. 2024b. So hateful! building a multi-label hate speech annotated arabic dataset. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15044–15055.

A Label distribution

B Examples from dataset

C Parameter Settings

For both Subtask 1 and Subtask 2, we adopted the DoRA-enhanced transformer model, "Gemma", and set the parameters as follows. The learning rate was set to 1×10^{-4} with no weight decay applied. The model was trained for 3

Text (Arabic)	Translation (English)	Label
@bka951753 هذا خطاب نفس الحوثي ...ثورجي من حق جيفارا .. حدث السوفتوير يا سقدي ..	@bka951753 This is the same rhetoric as the Houthis... a revolutionary like Che Guevara.. update your software, you fool..	hate
وجيه شرفت الكويت خير تشريف ????????????!	Wajih, you have honored Kuwait with the best representation ????????????!	hope
كسم الاولمبياد الناظر بيبدا علي اون تي في	Damn the Olympics, the headmaster starts on ON TV	not_applicable

Figure 3: Data example for subtask 1

Text (Arabic)	Translation (English)	Emotion
سيسي خاين..سيسي قتال #هتافات_ثورية	Sisi is a traitor.. Sisi is a killer #Revolutionary_chants	anger
من ينتزع ارواح اطفالنا من أجسادها بكل وحشية عليه ان يخشى ويخاف من غضبنا ويقلق من ردنا ويرتعد من بأس الله الذي منحنا إياه فـ صواريخنا المجنحة والبالستية و طائراتنا المسيرة ستلاحق المجرمين الى عقر دورهم وهذا ما سوف يحدث بلا شك بإذن الله #أطفال_اليمن	Whoever takes the souls of our children so brutally must fear our anger, worry about our response, and tremble before God's might given to us. Our missiles and drones will pursue the criminals into their homes, and this will surely happen, God willing. #Children_of_Yemen	anticipation
الوطني يخدم الوطن ويوثق تاريخه..	The patriot serves the homeland and records its history..	confidence
مش هنسمح بشويه فاسده ان يجيبوا سيره نادينا او يقربوا منه #ترك_يشغل_اعلام_يرقص	We will not allow a few corrupt people to mention our club or get close to it #Turki_shakes_media_dances	disgust
الخوف من الاتفاقيه والتطبيع ان اليهود هيسعوا فى بث الفتن بين دول الجوار ودى سياستهم)فرق تسد (وتمويل جماعه جديده زى الاخوانجيه لبث الفتن ونشر الارهاب على الارض فى الدول العربيه تماما كما حدث بعد اتفاقية السلام مع مصر فقد عملوا على إحياء الاخوانجيه مره اخرى بعد ما ناصر قضى عليهم.	The fear of the agreement and normalization is that the Jews will try to spread discord among neighboring countries (divide and rule), fund a new group like the Muslim Brotherhood to spread sedition and terrorism in Arab countries, just as they did after the peace agreement with Egypt when they revived the Brotherhood after Nasser eliminated them.	fear
ليك حق تضحك ياعمهم مانت فشختهم 😂	You're right to laugh, man, you destroyed them 😂	joy
حبيبي والله اكثثرر يارب ❤️❤️امين	My love, honestly even more, God willing, amen ❤️❤️	love
أحد التجار الشباب العمانيين يقول للأسف لما يكون عندهم كاش يروحوا هايبرماركت ولما يريدوا صبر يتسوقوا من عندي !!متى سندرك أن تسوقنا من تاجر عماني فتح لبيت عماني ودعما لإقتصاد الوطن ، وإذا اردتم التأكد فسألوا موظفي البنوك كم من آلاف الريالات يحولها التجار الأجانب إلى	A young Omani trader says: unfortunately, when they have cash, they go to the hypermarket, but when they want credit, they shop from me!! When will we realize that buying from an Omani trader opens an Omani home	neutral

Figure 4: Data example for emotion column of subtask 2

الخارج يوميا .	and supports the national economy? If you want proof, ask bank employees how many thousands of riyals foreign traders transfer abroad daily.	
مجموعه القدرة الجنسيه 👍 بديل الفياجرا والسيالس 💊 زيادة الانتصاب والصلابة 💊 علاج القذف السريع 💊 طبية وآمنه جداً لمرضى السكر والضغط 💊 معالجه البرود الجنسي 💊 زياده طول العضو والحجم 💊 للطلب والاستفسار 📲 للتواصل عبر الواتس د.لينا رمال	Sexual power package 👍 Alternative to Viagra and Cialis 💊 Increases erection and firmness 💊 Treats premature ejaculation 💊 Safe for diabetics and hypertensive patients 💊 Treats sexual coldness 💊 Increases length and size 💊 For orders and inquiries, contact via WhatsApp Dr. Lina Ramal 📲	optimism
عينكم على اخر مقطع 😂😂😂😂 انفخوا في زباله الدوري ما اوصيكم عنده ديفيز يفوز ٨٠ قلتوا لي مستعد احلف ما يفوز شيء هل المختنق	Keep your eyes on the last clip 😂😂😂😂 Blow up the trash of the league, I insist. He has Davies winning 80, but I told you I swear he won't win anything. Is he choking?	pessimism
نفسي ادخل جسمي أرتب عمودي الفقري و أزيته وأولع بخور وأشغل قرآن وأطلع	I wish I could enter my body, fix my spine, oil it, light incense, play the Quran, and leave.	sadness
ثواني بس انت هتقلل منه علشان البطوله ؟ !انت بتقول ميسي مبتقولش برشلونا ؟ !طيب والظاهره بتاعتك مجبتش البطوله مع اليوفي لي السنه دي ؟!	Wait a second, are you belittling him just because of the championship?! You're saying Messi, not Barcelona?! Well, your legend didn't win the championship with Juventus this year either?!	surprise

Figure 5: More example for emotion column of subtask 2

Table 5: Label distribution in the EmoHopeSpeech dataset (subtask2)

Column name	Label	Frequency
Emotion	Anger	1551
	Disgust	777
	Neutral	661
	Love	593
	Joy	533
	Anticipation	491
	Optimism	419
	Sadness	335
	Confidence	210
	Pessimism	194
	Surprise	143
	Fear	53
Offensive	No	4216
	Yes	1744
Hate (if Offensive = Yes)	Not_hate	1431
	Hate	303

Text (Arabic)	Translation (English)	Offensive
أحد التجار الشباب العمانيين يقول للأسف لما يكون عندهم كاش يروحوا هايبرماركت ولما يريدوا صبر يتسوقوا من عندي !!متى سندرك أن تسوقنا من تاجر عماني فتح لبيت عماني ودعما لإقتصاد الوطن ، وإذا اردتم التأكد فسألوا موظفي البنوك كم من آلاف الريالات يحولها التجار الأجانب إلى الخارج يوميا .	A young Omani trader says: unfortunately, when they have cash, they go to the hypermarket, but when they want credit, they shop from me!! When will we realize that buying from an Omani trader opens an Omani home and supports the national economy? If you want proof, ask bank employees how many thousands of riyals foreign traders transfer abroad daily.	no
مش هنسمح بشويه فاسده ان يجيبوا سيره نادينا او يقربوا منه #ترك_يشغل_اعلام_يرقص	We will not allow a few corrupt people to mention our club or get close to it #Turki shakes media dances	yes

Figure 6: Data example of offensive column of subtask 2

epochs, with a per-device training batch size of 1 and gradient accumulation over 4 steps to simulate a larger batch size. Warmup steps were set to 10, and the optimizer used was `paged_adamw_8bit`. We enabled mixed precision training with bf16 for efficiency. The maximum sequence length for tokenization was 1024, and padding was applied dynamically using the `DataCollatorWithPadding`.

Model checkpoints were saved every 50 steps,

Text (Arabic)	Translation (English)	Hate
مخصصه للاجانب فقط والسعودي كخه !!اشياء عجيبة غريبة ماتشوفها غير في السعودية !!المشكلة الاجانب نفسهم في دولهم مايعمل...	Reserved only for foreigners and the Saudi is worthless!! Strange and bizarre things we only see in Saudi Arabia!! The problem is that foreigners themselves don't do this in their own countries...	hate
مش هنسمح بشويه فاسده ان يجيبوا سيره نادينا او يقربوا منه #ترك_يشغل_اعلام_يرقص	We will not allow a few corrupt people to mention our club or get close to it #Turki shakes media dances	not_hate

Figure 7: Data example of hate column of subtask 2

and logging was performed every 10 steps. Unused columns in the dataset were removed to optimize memory usage. The DoRA configuration was applied with a rank r of 4, LoRA alpha of 32, LoRA dropout of 0.1, and targeting the projection layers `q_proj` and `v_proj`.

D Confusion Matrices

This appendix presents the confusion matrices for both subtasks, providing detailed visualization of the classification performance.

D.1 Subtask 1: Hate and Hope Speech Classification

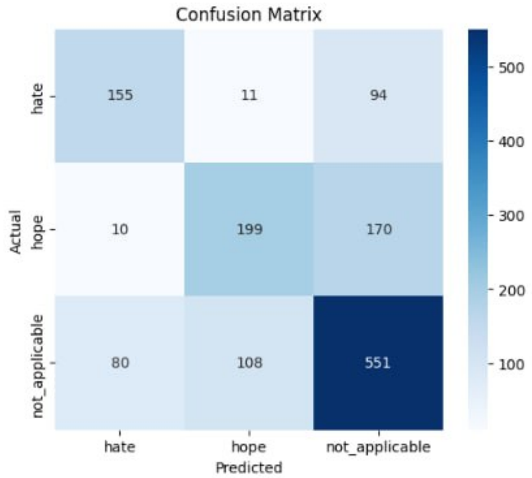

Figure 8: Confusion matrix for Subtask 1 (Hate and Hope Speech Classification) using Gemma-7B model.

D.2 Subtask 2: Multi-label Classification

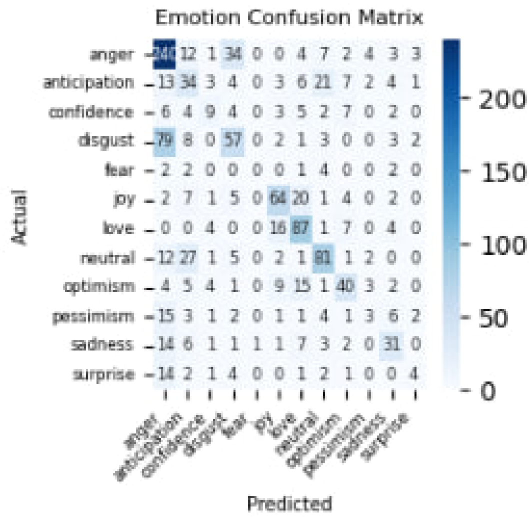

Figure 9: Confusion matrix for Emotion classification in Subtask 2.

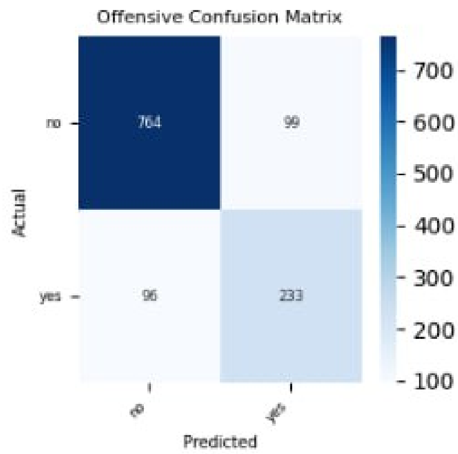

Figure 10: Confusion matrix for Offensive content detection in Subtask 2.

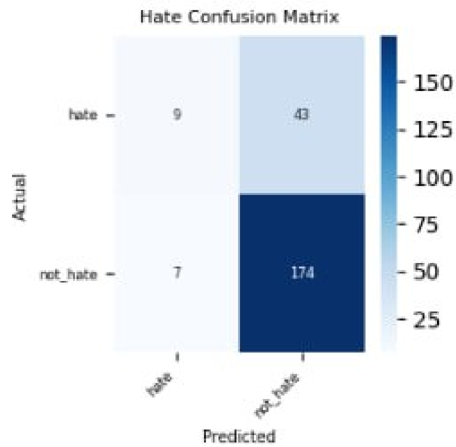

Figure 11: Confusion matrix for Hate speech detection in Subtask 2.