SYNAPSE@DravidianLangTech 2026: Multi-Level Political Meme Classification for Tamil and Malayalam

Suriya K P[1], Durai Singh K[1], Gnanasabesan G[1], Ganesh Sundhar[1], Hari Krishnan N[1], Jyothish Lal G[1]

[1]Amrita School of Artificial Intelligence, Coimbatore, Amrita Vishwa Vidyapeetham, India

{cb.en.u4aie22164, cb.en.u4aie22167, cb.en.u4aie22018, cb.en.u4aie22017 cb.en.u4aie22020}@cb.students.amrita.edu, g_jyothishlal@cb.amrita.edu

Abstract

Political memes in Tamil and Malayalam present unique multimodal challenges for automated understanding, combining visual context with code-mixed, culturally grounded text. We present **SYNAPSE**, our system for the DravidianLangTech@ACL 2026 shared task on multi-level political meme classification. The task requires hierarchical classification of memes along two levels: Level 1 identifies the political stance (Support/Praise vs. Troll/Oppose), and Level 2 identifies the target (individual person vs. party). Our approach fine-tunes the **Qwen3-VL-2B-Instruct** vision-language model using parameter-efficient LoRA adapters on task-specific multimodal data, with structured output prompting for hierarchical label prediction. We report results for both Tamil and Malayalam subtracks. For Malayalam, our system achieves a Level 1 F1 of 0.9200 and Level 2 F1 of 0.4256 (Avg-F1: 0.6728, Rank 5). For Tamil, our system achieves a Level 1 F1 of 0.7840 and Level 2 F1 of 0.4885 (Avg-F1: 0.6362, Rank 14).

Keywords: Political meme classification, multimodal NLP, vision-language models, Tamil, Malayalam, LoRA fine-tuning, DravidianLangTech.

1 Introduction

Political memes have become a potent instrument in online political discourse, particularly in multilingual and low-resource languages such as in Tamil and Malayalam social media. Unlike plain text, memes combine image and text in culturally situated ways that make their automated interpretation more challenging (Kiela et al., 2020). In the Tamil and Malayalam social media landscape, political memes routinely employ code-switching, sarcasm, regional symbols, and implicit references to political events that are vague to standard NLP systems (Francis et al., 2025).

The DravidianLangTech@ACL 2026 shared task on Multi-Level Political Meme Classification addresses this gap by making it a hierarchical classification challenge over Tamil and Malayalam political memes. Each meme must be classified along two levels simultaneously: Level 1 states whether the meme expresses *Support/Praise* or *Troll/Oppose* toward some political entity or person, and Level 2 identifies the target of that stance, specifically whether it is directed at an individual person or a party. The fine-grained nature of Level 2, with categories including *Troll Against Person*, *Troll Against Party*, *Support for Person*, and *Support for Party*, requires systems to go beyond surface level sentiment understading toward in detail political target identification.

We present SYNAPSE, a multimodal classification system based on parameter-efficient fine-tuning of the Qwen3-VL-2B-Instruct vision-language model (Qwen Team, 2025). Our system formats hierarchical classification as structured generation: a single pass of the model produces both Level 1 and Level 2 predictions via a constrained output template. This approach avoids the need for decoupled classification heads or two-stage inference pipelines, making it computationally efficient and straightforward to deploy.

Our key contributions are as follows:

1. **Multimodal fine-tuning** of a compact vision-language model (Qwen3-VL-2B) using LoRA adapters, enabling effective learning from small task-specific datasets without full model retraining.

2. **Structured output prompting** that defines joint Level 1 and Level 2 predictions in a single generation pass via a fixed template format, ensuring consistent hierarchical label extraction.

3. **Empirical comparison** of multiple ap-

Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 381–386

proaches including knowledge distillation from Qwen-8B and chain-of-thought prompting with translation, demonstrating that direct LoRA fine-tuning on the 2B model outperforms these alternatives on this task.

2 Related Work

2.1 Multimodal Meme Understanding

Meme understanding has received increasing attention in the NLP and computer vision research. Kiela et al. (2020) introduced the Hateful Memes dataset, establishing multimodal classification of meme content into a benchmark challenge and prooving that vision-language approach outperforms single modal baselines. Subsequent work has explored late fusion, attention mechanisms, and pre-trained vision-language models for meme understanding tasks (Pramanick et al., 2021).

2.2 Political Sentiment in Dravidian Languages

Sentiment analysis in Dravidian languages has been a recurring theme in the DravidianLangTech workshop series (Chakravarthi et al., 2022). Prior work has used transformer-based models including mBERT, IndicBERT, and XLM-RoBERTa for code-mixed Tamil and Malayalam text (Kakwani et al., 2020; Conneau et al., 2020). However, most prior systems treat the problem as purely text-based problem, neglecting the visual component of meme content.

2.3 Vision-Language Models for Low-Resource languages

Large-scale vision-language models (VLMs) such as LLaVA (Liu et al., 2023), InstructBLIP (Dai et al., 2023), (Kumar et al., 2024), (Premjith et al., 2023), and the Qwen-VL series (Bai et al., 2023) have demonstrated strong performance on image-text tasks with appropriate fine-tuning. Parameter-efficient fine-tuning methods such as LoRA (Hu et al., 2022) make it feasible to adapt large models to low-resource multimodal tasks without prohibitive computational cost. Our approach builds on this paradigm, applying LoRA fine-tuning to Qwen3-VL-2B-Instruct for politically in depth, Dravidian-language meme classification.

3 Task and Dataset Description

3.1 Task Definition

The shared task requires hierarchical classification of political memes along two levels:

- **Level 1:** Binary stance classification — *Support/Praise* vs. *Troll/Oppose*.
- **Level 2:** Target identification — whether the stance is directed at an *Individual Person* or a *Party* (or *Intersection* in the Malayalam track).

The full Level 2 label taxonomy across the two tracks is as follows. For Tamil: *Troll/Oppose Against Person*, *Troll/Oppose Against Party*, *Support for Person*, and *Support for Party*. For Malayalam: the corresponding categories additionally include *Intersection*, referring to memes that target an intersection of a person and a party simultaneously.

Evaluation uses macro-averaged precision, recall, and F1 score, computed independently for Level 1 and Level 2 predictions, and reported as per the scikit-learn classification report. The final ranking is based on the average of Level 1 and Level 2 macro F1 scores.

3.2 Datasets

Two separate datasets are provided for Tamil and Malayalam. Each meme is represented by an image file and a structured label file (Rajiakodi et al., 2026). Table 1 summarises partition sizes across both language tracks.

Language	Split	Instances	L2 Classes
Tamil	Train	803	4
	Test	201	4
Malayalam	Train	500	5
	Test	100	5

Table 1: Dataset statistics for Tamil and Malayalam tracks.

3.3 Class Distribution

Both datasets exhibit notable class imbalance. In the Tamil training set, *Troll/Oppose Against Person* constitutes the majority class (547 instances), followed by *Troll/Oppose Against Party* (146), *Support for Person* (86), and *Support for Party* (24). In the Malayalam training set, *Against Individual Person* is also dominant (315), while *Support for Party* and *Support for Individual Person* have far fewer

instances (10 and 12 respectively). This imbalance motivates the use of macro-averaged metrics and poses a significant challenge for minority-class performance at Level 2.

4 Methodology

Our system fine-tunes the Qwen3-VL-2B-Instruct vision-language model (Qwen Team, 2025) using Low-Rank Adaptation (LoRA) (Hu et al., 2022) to perform joint hierarchical classification across both levels in a single structured generation pass.

4.1 Base Model

We use **Qwen3-VL-2B-Instruct** as our backbone. This model provides a compact yet capable multimodal encoder-decoder architecture, jointly processing image and text inputs via a visual encoder and a transformer decoder. Its instruction-tuned variant supports structured dialogue-format inputs, making it well-suited for template-based output generation. Images are processed at a maximum resolution of 512×512 pixels to balance visual fidelity with computational efficiency.

4.2 Data Preprocessing

Training labels are normalised to a unified vocabulary prior to fine-tuning. Level 1 labels are mapped to `SUPPORT` and `CRITICISE`, while Level 2 labels are mapped to `PERSON`, `PARTY` and `BOTH`. Each training instance is converted into a structured conversation consisting of a user turn containing the meme image and a fixed instruction prompt, and an assistant turn containing the ground-truth structured output:

```
INTENT: [SUPPORT|CRITICISE]
TARGET: [PERSON|PARTY|BOTH]
```

This format enforces a predictable output structure that can be deterministically parsed at inference time, eliminating the ambiguity of free-form generation.

4.3 LoRA Fine-Tuning

We apply LoRA to the query and value projection matrices (`q_proj` and `v_proj`) of the transformer decoder. Fine-tuning hyperparameters are as follows: rank $r = 4$, $\alpha = 8$, dropout $= 0.15$. Training uses a per-device batch size of 1 with gradient accumulation over 8 steps (effective batch size 8), a learning rate of 5×10^{-5}, and a weight decay of 0.1. Models are trained for 500 steps with bfloat16 mixed precision. Only the LoRA adapter weights are saved after training, keeping the checkpoint lightweight and the base model frozen.

4.4 Inference and Output Parsing

At inference time, each test meme is passed through the model with the same instruction prompt used during fine-tuning. Generation is performed greedily (`do_sample=False`) with a maximum of 150 new tokens. The output is parsed line-by-line: the `INTENT:` line yields the Level 1 prediction and the `TARGET:` line yields the Level 2 prediction. Unknown or malformed outputs are assigned a fallback label of `UNKNOWN`.

4.5 Explored Alternatives

We investigated two additional approaches that were ultimately outperformed by direct LoRA fine-tuning:

- **Knowledge distillation from Qwen3-VL-8B:** We attempted to use the 8B variant as a teacher model to generate soft labels or explanations for training the 2B student. While the 8B model produced higher-quality zero-shot outputs, the distillation process did not yield consistent improvements over direct task fine-tuning of the 2B model, likely due to domain mismatch and the small training set size.

- **Chain-of-thought prompting with translation:** We explored prompting the model to first translate the Tamil or Malayalam meme text to English and then reason step-by-step before predicting the final labels. This approach consistently underperformed, as translation introduced errors on code-mixed and colloquial inputs, and the additional reasoning steps were not reliably grounded in the meme's visual content.

These comparisons confirm that direct, targeted fine-tuning on the task data with structured output constraints is the most effective strategy for this low-resource multimodal classification problem.

5 Results and Analysis

5.1 Evaluation Metric

Performance is reported using macro-averaged precision (P), recall (R), and F1 score separately for Level 1 and Level 2 predictions. Accuracy (ACC) is also reported. The final ranking metric is the

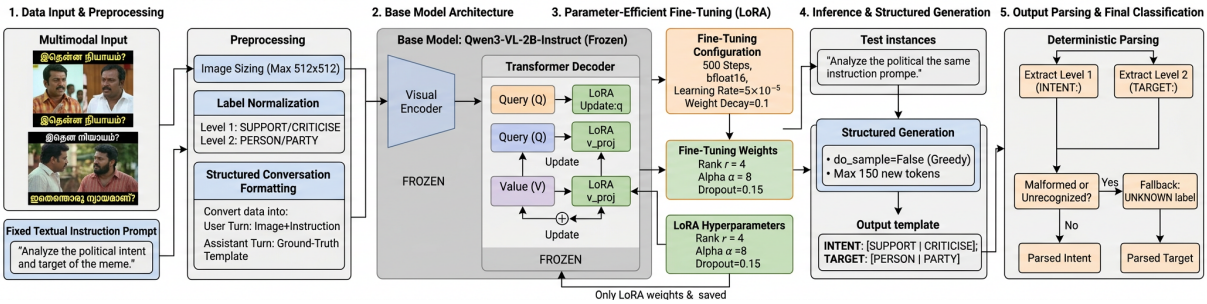

Figure 1: Overview of the proposed SYNAPSE architecture for multi-level political meme classification. The pipeline processes each meme through parallel image and text streams, feeds them into a Qwen3-VL-2B-Instruct vision-language model fine-tuned with LoRA adapters on query/value projections (base model frozen), and produces hierarchical predictions via structured output prompting: Level 1 stance (Support vs. Troll) and Level 2 target (Person vs. Party).

average of Level 1 and Level 2 macro F1 scores:

$$\text{Avg-F1} = \frac{\text{F1}_{\text{Level1}} + \text{F1}_{\text{Level2}}}{2} \quad (1)$$

5.2 Official Results

Tables 2 and 3 report our official shared task results alongside the top-ranked systems for the Malayalam and Tamil subtracks respectively.

Team	L1-F1	L2-F1	Avg-F1	Rank
SYNAPSE	**0.9200**	**0.4256**	**0.6728**	**5**

Table 2: Official results - Malayalam subtrack (selected teams).

Team	L1-F1	L2-F1	Avg-F1	Rank
SYNAPSE	**0.7840**	**0.4885**	**0.6362**	**14**

Table 3: Official results - Tamil subtrack (selected teams).

5.3 Analysis

Across both subtracks, Level 1 performance substantially exceeds Level 2 performance. This is expected: binary stance classification (Support vs. Troll) is a simpler task with higher inter-class separability, whereas Level 2 target identification requires resolving fine-grained distinctions between person- and party-directed stances that often co-occur in the same meme.

The Malayalam subtrack achieved a higher Level 1 F1 (0.9200) than Tamil (0.7840), possibly reflecting a cleaner visual signal in the Malayalam memes or a more balanced training distribution at Level 1. The Tamil Level 2 F1 (0.4885), however, slightly exceeds the Malayalam Level 2 F1 (0.4256), suggesting that the Tamil training data provides somewhat stronger supervision for fine-grained target identification, despite its severe class imbalance at the minority categories.

The low Level 2 scores across all submitted systems reflect the inherent difficulty of this sub-problem: distinguishing person-directed from party-directed political stance in memes requires the system to recognize political figures and party symbols from visual content, a capability that compact VLMs with limited fine-tuning data may struggle to acquire reliably. Future work could address this by incorporating entity recognition or political knowledge grounding.

6 Conclusion

We presented SYNAPSE, a LoRA-fine-tuned vision-language system for multi-level political meme classification in Tamil and Malayalam. By framing hierarchical classification as structured text generation from a multimodal input, our system avoids the complexity of multi-head architectures while achieving competitive performance. Our system ranked 5th on the Malayalam subtrack (Avg-F1: 0.6728) and 14th on the Tamil subtrack (Avg-F1: 0.6362).

Comparative experiments showed that direct LoRA fine-tuning on task data outperforms knowledge distillation from a larger model and chain-of-thought prompting with translation, highlighting the importance of task-specific adaptation over general reasoning strategies for this low-resource multimodal domain.

Code: https://github.com/SURIYA-KP/Poli-Meme

Limitations

1. **Small training sets:** With 803 Tamil and 500 Malayalam training instances, both datasets are small for multimodal fine-tuning. Minority classes at Level 2 have very few examples, directly limiting classification performance on those categories (Francis et al., 2026).

2. **Label space inconsistency across languages:** The Malayalam dataset includes an additional *Intersection* class at Level 2, but the Tamil dataset doesn't contain this class. Due to this mismatch in label taxonomies, we had to train separate models for each language, for the best performance instead of a unified model over the combined dataset, limiting its generalization.

3. **Visual grounding of political entities:** The system has no explicit mechanism to identify which political figure or party appears in a meme's image. Level 2 predictions rely on the model's implicit visual knowledge acquired during pretraining, which may be inconsistent for regional political figures less represented in the pretraining corpus.

4. **Resolution constraint:** Capping image resolution at 512×512 pixels may discard fine-grained textual content in memes where embedded text is small or dense, which is common in Tamil political memes.

Acknowledgments

We thank the organizers of DravidianLangTech@ACL 2026 for curating the Tamil and Malayalam political meme datasets and hosting the shared task. We also gratefully acknowledge the computational resources provided by our institution.

References

Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. 2023. Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond. *arXiv preprint arXiv:2308.12966*.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, Elizabeth Sherly, and 1 others. 2022. Overview of the shared task on sentiment analysis for dravidian languages in code-mixed text. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pages 1–9.

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 ACL*, pages 8440–8451.

Wenliang Dai, Junnan Li, Dongxu Li, Anthony Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale Fung, and Steven Hoi. 2023. Instructblip: Towards general-purpose vision-language models with instruction tuning. In *Advances in Neural Information Processing Systems*, volume 36.

Meclin A. Francis, Ayswarya R. Kurup, B. Premjith, and Bharathi Raja Chakravarthi. 2025. Multimodal fake news classification in tamil using fact-checked social media content and cost-sensitive learning. *IEEE Access*, 13:157477–157510.

Meclin A. Francis, Ayswarya R. Kurup, B. Premjith, Bharathi Raja Chakravarthi, and Saranya Rajiakodi. 2026. Tamilfacts: A comprehensive multimodal dataset of fact-checked social media content in tamil language. In *Speech and Language Technologies for Low-Resource Languages*, pages 167–182. Springer.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *Proceedings of ICLR*.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N C, Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. Indicnlpsuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for indian languages. In *Findings of EMNLP*, pages 4948–4961.

Douwe Kiela, Hamed Firooz, Aravind Natesan Rau, Virginie Testuggine, Alyssa Mensch, Devvret Mahajan, Josh Meyer, Maximilian Nickel, Dhruv Mahajan, and Laurens van der Maaten. 2020. The hateful memes challenge: Detecting hate speech in multimodal memes. In *Advances in Neural Information Processing Systems*, volume 33, pages 2611–2624.

MR Dinesh Kumar, Pillalamarri Akshaya, R Saivarsha, NT Shrish Surya, B Premjith, V Sowmya, and G Jyothish Lal. 2024. Enhanced vision language model for visual question answering in medical images. In *International Conference on Data Science and Applications*, pages 225–236. Springer.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. In *Advances in Neural Information Processing Systems*, volume 36.

Shraman Pramanick, Shivam Sharma, Dimitar Dimitrov, Md. Shad Akhtar, Preslav Nakov, and Tanmoy

Chakraborty. 2021. Momenta: A multimodal framework for detecting harmful memes and their targets. In *Findings of EMNLP*, pages 4439–4455.

B Premjith, G Jyothish Lal, V Sowmya, Bharathi Raja Chakravarthi, Rajeswari Natarajan, K Nandhini, Abirami Murugappan, Bharathi B, M Kaushik, Prasanth Sn, Aswin Raj R, and Vijai Simmon S. 2023. Findings of the shared task on multimodal abusive language detection and sentiment analysis in tamil and malayalam. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 72–79.

Qwen Team. 2025. Qwen3-vl technical report. *arXiv e-prints*.

Saranya Rajiakodi, Shunmuga Priya Muthusamy Chinnan, Premjith B, Subalalitha CN, Rahul Ponnusamy, Anshid K A, Bhuvaneshwari Sivagnanam, Jananayagan V, Bharathi Raja Chakravarthi, Ragavan N, and Santhini P. 2026. Overview of the shared task on multilevel political meme classification in tamil and malayalam. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.