Findings of the Shared Task on Misogyny Meme Detection: DravidianLangTech@NAACL 2025

Bharathi Raja Chakravarthi[1], **Rahul Ponnusamy**[2], **Saranya Rajiakodi**[3],
Shunmuga Priya Muthusamy Chinnan[1], **Paul Buitelaar**[2], **Bhuvaneswari Sivagnanam**[3],
Anshid Kizhakkeparambil[4]
[1]School of Computer Science, University of Galway, Ireland
[2]Data Science Institute, University of Galway, Ireland
[3]Department of Computer Science, Central University of Tamil Nadu, India
[4]WMO Imam Gazzali Arts and Science College, Kerala, India

Abstract

The rapid expansion of social media has facilitated communication but also enabled the spread of misogynistic memes, reinforcing gender stereotypes and toxic online environments. Detecting such content is challenging due to the multimodal nature of memes, where meaning emerges from the interplay of text and images. The Misogyny Meme Detection shared task at DravidianLangTech@NAACL 2025 focused on Tamil and Malayalam, encouraging the development of multimodal approaches. With 114 teams registered and 23 submitting predictions, participants leveraged various pretrained language models and vision models through fusion techniques. The best models achieved high macro F1 scores (0.83682 for Tamil, 0.87631 for Malayalam), highlighting the effectiveness of multimodal learning. Despite these advances, challenges such as bias in the data set, class imbalance, and cultural variations persist. Future research should refine multimodal detection methods to improve accuracy and adaptability, fostering safer and more inclusive online spaces.

Disclaimer: This research paper contains offensive/harmful content for research purposes. Viewer discretion is advised.

1 Introduction

The widespread adoption of social media has transformed digital communication, allowing instantaneous sharing of ideas and fostering global connectivity (Singh et al., 2023). However, this evolution has also led to challenges, particularly the proliferation of harmful content such as misogyny memes (Gasparini et al., 2022). These memes, often combining visual and textual elements, propagate gender-based discrimination, perpetuate stereotypes, and contribute to toxic online environments. Their multimodal nature poses significant challenges for automated detection, as the nuanced interplay between images and text frequently conveys implicit and context-dependent meanings (Kumari et al., 2024). Traditional unimodal detection methods often fail to address this complexity, underscoring the need for advanced multimodal analysis techniques to effectively identify and mitigate such content.

Addressing misogyny in online spaces is a critical step toward fostering inclusive and respectful digital environments. Misogyny memes, by embedding discriminatory messages in humor or satire, not only normalize toxic behaviors but also marginalize women and reinforce societal inequalities. Detecting and moderating these memes is particularly challenging in low-resource contexts where annotated datasets and linguistic resources are scarce (Huang et al., 2024). The task requires innovative approaches that integrate textual and visual modalities to capture the implicit biases and indirect messaging characteristic of these memes. By advancing research in this area, it is possible to combat the spread of harmful content and contribute to safer and more equitable online spaces (Rizzi et al., 2024).

To address this issue, we conducted the second shared task on "Misogyny Meme Detection"[1] under the DravidianLangTech@NAACL 2025[23] initiative. This shared task focuses on the automatic detection of misogyny in memes across two languages, including Tamil and Malayalam which are low-resourced. The aim is to foster the development of computational models capable of identifying misogynistic content while accounting for linguistic and cultural variations in online communication.

The task is grounded in several objectives:

1. To encourage the creation of state-of-the-art

[1] https://codalab.lisn.upsaclay.fr/competitions/20856

[2] https://sites.google.com/view/dravidianlangtech-2025/

[3] https://2025.naacl.org/

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 86–96
May 3, 2025 ©2025 Association for Computational Linguistics

systems for misogyny detection in memes using multimodal approaches.

2. To promote research in low-resource languages, extending the applicability of NLP technologies beyond high-resource settings.

The shared task attracted significant participation from researchers around the world, demonstrating the growing recognition of the need to address misogyny in online spaces. A total of 114 teams registered for the shared task, with 23 teams successfully submitting their predictions, showcasing a diverse range of methodologies that used both textual and visual modalities to enhance multimodal classification performance. The results indicated that the team DLRG_RR achieved the highest macro F1-score (0.83682) for Tamil, while CUET_Novice team obtained the top macro F1-score (0.87631) for Malayalam, emphasizing the effectiveness of multimodal learning. Despite the success of fusion-based models, challenges such as class imbalance, dataset biases, and subtle variations in misogynistic language remain areas for further exploration. Future research should focus on mitigating these biases, addressing cultural nuances, and improving context-dependent understanding to improve the robustness of misogyny meme detection models, contributing to a safer and more inclusive digital space.

2 Related Work

2.1 Misogyny detection

Misogyny detection in online platforms has been a focal point of research, particularly as the internet continues to foster gender-based hate speech. Early efforts include the Evalita 2018 and IberEval 2018 shared tasks, which introduced the Automatic Misogyny Identification (AMI) challenge to detect misogynistic content in English and Italian texts (Fersini et al., 2018). SemEval 2019 extended this focus to multilingual hate speech, addressing misogyny alongside other forms of hate targeting immigrants and emphasizing the detection of aggressive and non-aggressive speech (Basile et al., 2019).

Recent advances have embraced transformer-based models such as BERT and RoBERTa for misogyny detection, leveraging their capability to understand nuanced language. Multilingual models have been particularly effective for tasks involving diverse linguistic contexts (Devlin et al., 2019; Liu et al., 2019). In the multimodal space, datasets such as Facebook Hateful Memes and new multimodal misogyny-specific datasets have encouraged combining textual and visual cues, as seen in approaches utilizing VisualBERT, UNITER, and CLIP for better classification accuracy (Chen et al., 2020; Radford et al., 2021).

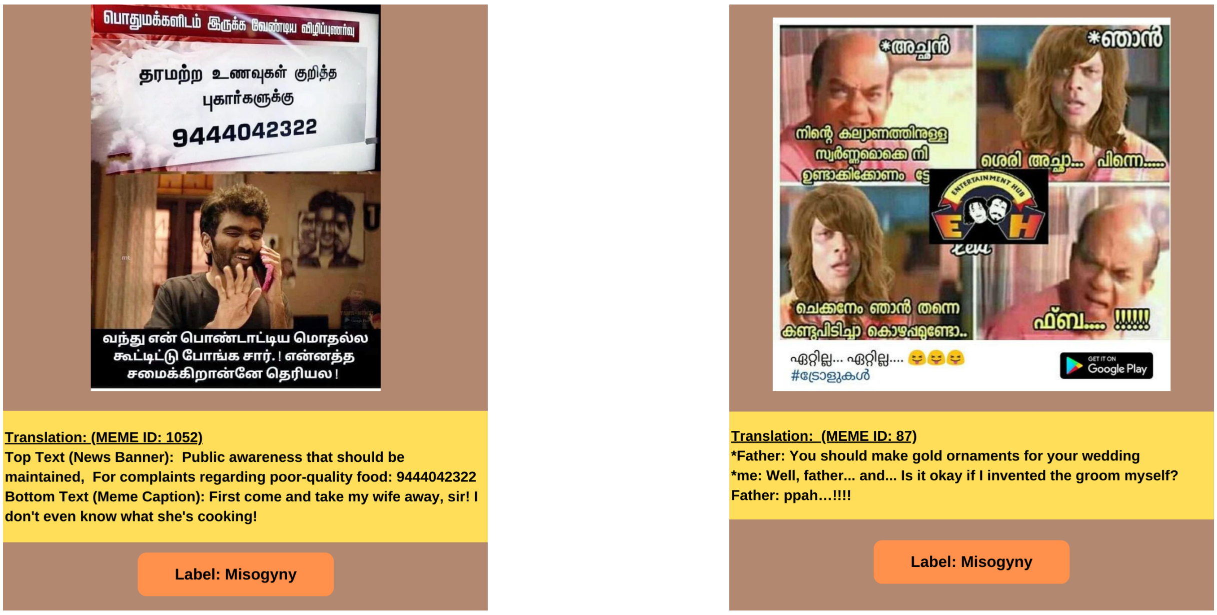
2.2 Multimodal Classification

Multimodal classification methods have become pivotal in tackling tasks involving combined visual and textual data. Traditional approaches used separate pipelines for image and text processing, combining the outputs through simple fusion techniques. Suryawanshi et al. (2020) explored an early fusion technique that combines textual and visual modalities at the embedding level, demonstrating its effectiveness compared to unimodal baselines focused solely on text or images. Koutlis et al. (2023) introduced MemeFier, a deep learning-based framework featuring a dual-stage modality fusion module. This system captured intricate inter-modal connections by integrating feature-level alignment with token-level modality interactions, achieving state-of-the-art results in fine-grained meme classification.

2.3 Related shared tasks

Numerous shared tasks have significantly advanced research on misogyny detection and multimodal content analysis. The Evalita 2018 and IberEval 2018 shared tasks introduced the Automatic Misogyny Identification (AMI) challenge, focusing on detecting misogynistic content in English and Italian (Fersini et al., 2018). These tasks were among the first to address gender-based hate speech systematically. SemEval 2019 expanded the focus to multilingual hate speech detection, including misogyny and hate against immigrants, with distinctions between aggressive and non-aggressive content in English and Spanish (Basile et al., 2019).

Multimodal shared tasks have further enriched the research landscape. The Memotion shared tasks (2020, 2022) targeted the sentiment and emotion analysis of memes, providing valuable benchmarks for multimodal sentiment classification (Sharma et al., 2020; Patwa et al., 2022). The MultiOFF dataset for offensive meme detection highlighted challenges in integrating text and visual modalities for hate speech classification (Suryawanshi et al., 2020). TrollsWithOpinion (2023) introduced a three-level taxonomy for trolling and opinion ma-


பொதுமக்களிடம் இருக்க வேண்டிய விழிப்புணர்வு
தரமற்ற உணவுகள் குறித்த புகார்களுக்கு
9444042322
வந்து என் பொண்டாட்டிய மொதல்ல கூட்டிட்டு போங்க சார்.! என்னத்த சமைக்கிறான்னே தெரியல!
*അച്ഛൻ
*ഞാൻ
നിന്റെ കല്യാണത്തിനുള്ള സ്വർണ്ണമൊക്കെ നീ ഉണ്ടാക്കിക്കോണം ട്ടേ
ശെരി അച്ഛാ... പിന്നെ.....
ചെക്കനേം ഞാൻ തന്നെ കണ്ടുപിടിച്ചാ കൊഴപ്പുണ്ടോ..
ഫ്ബ.... !!!!!
ഏറ്റില്ല... ഏറ്റില്ല.... 😂😂😂
#ട്രോളുകൾ
GET IT ON Google Play

Translation: (MEME ID: 1052)
Top Text (News Banner): Public awareness that should be maintained, For complaints regarding poor-quality food: 9444042322
Bottom Text (Meme Caption): First come and take my wife away, sir! I don't even know what she's cooking!

Label: Misogyny

(a) Misogyny example from Tamil set

Translation: (MEME ID: 87)
*Father: You should make gold ornaments for your wedding
*me: Well, father... and... Is it okay if I invented the groom myself?
Father: ppah...!!!!

Label: Misogyny

(b) Misogyny example from Malayalam set

Figure 1: Examples from the Tamil and Malayalam sets

nipulation, emphasizing domain-specific opinion manipulation in memes (Suryawanshi et al., 2023).

Our previous shared task, organized as part of LT-EDI@EACL 2024, focused on multitask meme classification with an emphasis on identifying misogynistic and troll content in memes, specifically in Tamil and Malayalam. This task received significant participation, with 52 teams registering and notable submissions. The task A is focused on misogyny meme detection in Tamil and Malayalam language where the top method submitted by MUCS team (Mahesh et al., 2024) achieving macro F1 scores of 0.73 (Tamil) and 0.87 (Malayalam). This effort demonstrated the importance of regional language datasets and the effectiveness of multilingual computational approaches in tackling misogynistic meme detection (Chakravarthi et al., 2024).

3 Task Description

The Shared Task on Misogyny Meme Detection, organized as part of DravidianLangTech@NAACL 2025, aimed to challenge participants to develop advanced multimodal machine learning systems capable of analyzing both textual and visual components of memes. The primary objective of the task was to classify memes as either Misogyny or Non-Misogyny in Tamil and Malayalam languages (Ponnusamy et al., 2024), emphasizing the nuanced intersection of multilingualism and multimodality in the analysis of social media content. The examples from the dataset is show in the Figure 1.

Participants were initially provided with training and development datasets to build and validate their models. Subsequently, a test dataset without labels was released for the final evaluation of the models, which were trained on the previously provided datasets. Submissions was required in a predefined CSV format with their model's prediction using the test set provided for the evaluation. After the release of the results, the labeled test set was shared with participants for personal verification and further analysis.

The datasets included an image folder containing memes in JPG format, accompanied by a CSV file comprising image_id, labels (indicating misogyny or non-misogyny), and transcriptions (text extracted from memes). Participants were expected to adhere to strict submission guidelines and provide their predictions in a predefined format. This shared task served as a platform for exploring cutting-edge methodologies in Natural Language Processing (NLP) and multimodal learning, driving innovation in the analysis of multilingual and multimodal social media content.

4 Participant's Methods

For this shared task, there are total of 114 teams registered, among them we have received a total of 23 submissions where all the participants have used various types of methodologies to address the task of detecting misogyny and non-misogyny memes in Tamil and Malayalam languages in a multimodal settings from the dataset provided. The participants rank list has been mentioned in the Table 1 for Tamil language and Table 2 for Malayalam language.

- **byteSizedLLM** (Manukonda and Kodali, 2025): This team proposed a multimodal approach for misogynistic meme detection by combining textual and visual features. They fine-tuned the XLM-RoBERTa Base model on Tamil and Malayalam text from the AI4Bharath dataset, using IndicTrans to generate native script, Romanized, and partially transliterated text variations. Text embeddings were mapped to a 768-dimensional space. To align with this, they modified ResNet-50's fully connected layer to extract image features in the same dimension. An Attention-Driven BiLSTM was used to fuse the modalities, capturing sequential dependencies. An attention mechanism followed by a fully connected layer optimized with cross-entropy loss enabled accurate classification.
- **CUET_NetworkSociety**: This team extracted visual features using data augmentation techniques to enhance generalization, employing ResNet18 and EfficientNet-B4. For text processing, they used transformer-based models including IndicBERT for Indian languages, LaBSE for multilingual embeddings, and XLM-RoBERTa for contextual understanding. They applied both classical classifiers—Logistic Regression, SVM, and Random Forest—and deep learning models such as CNN, BiLSTM, and BiLSTM+CNN for textual classification. To integrate text and image modalities, they experimented with concatenation-based fusion (feature-level) and late fusion (prediction-level).
- **Dll5143** (Pattanaik et al., 2025): This team utilized three models—M-CLIP, IndicBERT, and Google's MuRIL—each fine-tuned separately on Tamil and Malayalam meme data to detect misogyny from both images and text. The individual model predictions were combined using majority voting, which enhanced robustness and accuracy by capturing diverse features across modalities and languages. This ensemble approach effectively fused insights from distinct multilingual and multimodal models, improving detection accuracy for memes in Tamil and Malayalam.
- **CUET-NLP_Big_O** (Hossan et al., 2025): This team resized images to 224×224 pixels and normalized them using ImageNet statistics. Tamil and Malayalam texts were tokenized with MuRIL (128 tokens). Visual features were extracted using DenseNet121, EfficientNetB0, ResNet50, and VGG19; textual features used TF-IDF, BoW, and 100-dimensional GloVe embeddings. Features were fused via a fully connected classifier. Training ran for 45 epochs with a 2e-5 learning rate, batch size of 16, and ReduceLROnPlateau scheduler on dual NVIDIA Tesla T4 GPUs, taking 120–150 minutes.
- **Code_Conquerors** (Rao et al., 2025): This team used CLIP model embeddings for image features and BERT-base uncased embeddings for text. For Tamil data, they trained a hybrid model combining ResNet for images and BERT-base uncased for text. For Malayalam data, Vision Transformer replaced ResNet, paired again with BERT-base uncased. In both cases, image and text embeddings were concatenated before training, allowing the model to effectively learn and integrate visual and textual context for improved misogyny detection.
- **Shraddha**: Here, text features were extracted using BiLSTM, and attention mechanisms, while image features were obtained using the ImageNet pre-trained MobileNetV2 model with attention layers. These features were fused for classification, optimized with focal loss and class weights to handle class imbalance.
- **LexiLogic** (M et al., 2025): This study uses L3Cube-Malayalam-BERT and L3Cube-Tamil-BERT for meme categorization and abusive language detection. Data is preprocessed through tokenization and normalization. Fine-tuning uses cross-entropy loss over five epochs at a 2e-5 learning rate. Language-specific embeddings and data augmentation improve performance on low-resource Indian

languages, effectively handling hostile language.

- **One_by_zero** (Chakraborty et al., 2025): They employed CNN, VGG16, and Vision Transformer (ViT) models for visual features extraction, optimizing ViT using the AdamW optimizer and binary cross-entropy loss. BiLSTM, TextCNN, LSTM+CNN, Malayalam BERT, and IndicBERT were used to extract textual features; they were all trained using the Adam optimizer and binary cross-entropy loss. These features were concatenated at a fusion layer and then run through a fully connected classifier that was tuned with Adam and binary cross-entropy loss.

- **teamiic** (Sharma et al., 2025): The XLM-R model was used to handle text data, and the Vision Transformer (ViT) was used to extract features from images. To create a single representation, the embeddings from the two modalities were concatenated. For classification, a proprietary neural network classifier with ReLU activation and a fully connected hidden layer was employed along with cross Entropy Loss and the Adam optimizer.

- **Team_Strikers** (Shanmugavadivel et al., 2025a): This team used LSTM and GRU models to process Tamil-English code-mixed text with TF-IDF, GloVe, and Word2Vec embeddings, while ResNet and EfficientNet CNNs extracted visual features. The CNN-LSTM model combined spatial and sequential learning. Despite challenges with code-mixed input in ResNet-BERT, the GRU-EfficientNet model effectively merged text and visuals.

- **CUET-823** (Mallik et al., 2025): For both text and image inputs, they used text-based augmentation (back-translation via Tamil-English-Tamil and Tamil-Malayalam-Tamil) and image alterations (brightness adjustment, grayscale, posterization) to address class imbalance. They experimented with ViT, ResNet, and EfficientNet for pictures, training for 20 epochs (batch size: 16, learning rate: 1e-4), and optimized mBERT and IndicBERT for text with a 512-token length using the AdamW optimizer (learning rate: 2e-5).

- **CUET_Novice** (Sayma et al., 2025): This team developed a multimodal approach to detect misogyny in Malayalam memes by combining visual and textual features. They used an 8-layer CNN, ResNet-50, Vision Transformer (ViT), and Swin Transformer for visual feature extraction. Text was processed using Malayalam-BERT, generating 768-dimensional embeddings. These were fused with 1024-dimensional Swin Transformer features to create a 1792-dimensional vector. A two-layer neural network with ReLU activation and 0.1 dropout was used for classification. The team trained their model with the AdamW optimizer, binary cross-entropy loss, gradient clipping, a batch size of 16, and a 5e-5 learning rate over five epochs.

- **InnovationEngineers** (Shanmugavadivel et al., 2025b): This team applied padding to text sequences up to a length of 100 before using BERT to extract semantic features. For visual processing, images were resized to 224×224 pixels, normalized to [0, 1], and processed in batches using EfficientNetB0 for feature extraction. They combined both textual and visual features using a Vision-Language Model (VLM) for classification, effectively integrating multimodal data to enhance performance in misogynistic meme detection.

- **Zero_knowledge**: This team employed a multimodal approach, extracting image features using a Convolutional Neural Network (CNN) and processing text through an embedding layer followed by an Long Short-Term Memory (LSTM) to capture sequential and contextual information. Outputs from both branches were concatenated in a fusion layer to integrate visual and textual data. A fully connected dense layer refined the fused features, followed by a sigmoid activation for binary classification. The model was trained using the Adam optimizer with binary cross-entropy loss, ensuring stable and effective convergence for misogynistic meme detection.

- **LinguAIsts** (Arthir et al., 2025): This team used a Support Vector Machine (SVM) with a linear kernel for binary classification, leveraging its effectiveness with textual input vectorized using TF-IDF. To optimize performance, they applied GridSearchCV to tune hyperparameters such as C, kernel, and gamma, using five-fold cross-validation for robust model selection.

- **Fired_from_NLP** (Chowdhury et al., 2025): This team implemented a multimodal approach, extracting visual features using CNN models like EfficientNetB7, ResNet50, and MobileNetV2, and processing text with Tamil-BERT and Malayalam-BERT. Text preprocessing included padding, truncation, and attention masking via the BERT tokenizer. Images were resized to 224×224 pixels and standardized. Transformer models (mBERT, Indic-BERT, Tamil-BERT, Malayalam-BERT) handled textual feature extraction, while cross-modal attention was used to compute attention scores between text and image features. Outputs were fused using concatenation and the Hadamard product, then passed through dense layers for binary classification. The model was trained with binary cross-entropy loss, early stopping, and a learning rate scheduler.

- **Magma**: For this shared task, They utilized Google's Gemini model to generate dense vector embeddings ('models/embedding-001') which capture the semantic features of Malayalam text. These embeddings were then processed through a Random Forest classifier with 100 estimators, trained on an 80-20 train-test split. For Tamil text classification, They employed a BERT (Bidirectional Encoder Representations from Transformers) model, fine-tuning it specifically for Tamil language processing. The BERT model's bidirectional self-attention mechanisms were adapted to understand Tamil linguistic patterns through the fine-tuning process.

- **CUET-NLP_MP** (Mohiuddin et al., 2025): This team classified Tamil and Malayalam memes using both unimodal and multimodal models. For text, they tested CNN, SVM, Bi-LSTM, mBERT, and XLM-R; for images, they used VGG16, VGG19, ResNet50, Vision Transformer (ViT), and Swin Transformer. The best models for each language and modality were combined for multimodal analysis. Their final model integrated IndicBERT for text and ViT-Base-Patch16-224 for images, with fused embeddings passed through a dense classification layer. Trained over five epochs with a batch size of 16 and a learning rate of 2e-5, this multimodal setup delivered the team's best overall performance.

- **HerWILL** (Preeti et al., 2025): This team adopted a multimodal approach, using language-specific pre-trained models for text encoding: hate-speech-CNERG/malayalam-codemixed-abusive-MuRIL for Malayalam and Tamil-codemixed-abusive-MuRIL for Tamil. For visual features, they used OpenAI's CLIP model (openai/clip-vit-base-patch32) alongside an MLP classifier. They also experimented with a larger vision model (zer0int/CLIP-GmP-ViT-L-14) to explore performance gains. Both early and late fusion strategies were evaluated, along with language models like ai4bharat/IndicBERTv2MLM-only and PosteriorAI/dravida_llama2_7b.

- **vemuri_monisha**: This team combined both the image and text features for this classification task. The image features were extracted using Vision Transformer (ViT), while text features were derived from BERT. These features are then fused and passed through a Random Forest Classifier.

- **SemanticCuetSync**: They fine-tuned the large vision models such as LLaMa 3.2 vision 11b to detect the misogyny content in the dataset provided.

- **DLRG_RR**: This team have utilized the mBERT model to improve the contextual understand in both the Tamil and Malayalam languages.

- **MNLP** (Chauhan and Kumar, 2025): This team used XML-RoBERTa and Byte-Pair Encoding for the extraction of textual features and ViT for the visual features extraction. Then the concatenation based fusion mechanism has been applied and ML models like KNN, SVM, RF, NB and DL models such as LSTM, GRU and Multimodal classifier were used for classification task along with ReLU activation.

5 Results and Discussions

Participants predictions were collected in csv format and evaluated using the macro F1-score, a ro-

Table 1: Rank List of Tamil Language

Team Name	Run	F1 Score	RANK
DLRG_RR	1	0.83682	1
CUET-NLP_Big_O (Hossan et al., 2025)	3	0.81716	2
byteSizedLLM (Manukonda and Kodali, 2025)	3	0.80809	3
CUET-823 (Mallik et al., 2025)	3	0.78120	4
Dll5143 (Pattanaik et al., 2025)	2	0.77591	5
CUET-NLP_MP (Mohiuddin et al., 2025)	1	0.77180	6
CUET_NetworkSociety	1	0.76323	7
MNLP (Chauhan and Kumar, 2025)	2	0.73516	8
LinguAIsts (Arthir et al., 2025)	-	0.71259	9
Shraddha	1	0.70501	10
teamiic (Sharma et al., 2025)	-	0.68830	11
InnovationEngineers (Shanmugavadivel et al., 2025b)	2	0.68782	12
LexiLogic (M et al., 2025)	1	0.68707	13
Fired_from_NLP (Chowdhury et al., 2025)	1	0.67754	14
Code_Conquerors (Rao et al., 2025)	1	0.66142	15
Magma	1	0.65068	16
Team_Strikers (Shanmugavadivel et al., 2025a)	1	0.64776	17
Zero_knowledge	-	0.47801	18
SemanticCuetSync	1	0.40692	19

Table 2: Rank List of Malayalam Language

Team Name	Run	F1 Score	RANK
CUET_Novice (Sayma et al., 2025)	3	0.87631	1
HerWILL (Preeti et al., 2025)	1	0.87483	2
One_by_zero (Chakraborty et al., 2025)	3	0.86658	3
Dll5143 (Pattanaik et al., 2025)	2	0.84927	4
MNLP (Chauhan and Kumar, 2025)	1	0.84237	5
CUET-NLP_MP (Mohiuddin et al., 2025)	1	0.84118	6
teamiic (Sharma et al., 2025)	1	0.84066	7
byteSizedLLM (Manukonda and Kodali, 2025)	1	0.83912	8
CUET-NLP_Big_O (Hossan et al., 2025)	1	0.82531	9
LexiLogic (M et al., 2025)	1	0.80364	10
Fired_from_NLP (Chowdhury et al., 2025)	1	0.80347	11
CUET_NetworkSociety	1	0.80347	12
Code_Conquerors (Rao et al., 2025)	1	0.75649	13
Shraddha	1	0.75467	14
LinguAIsts (Arthir et al., 2025)	-	0.68186	15
Magma	1	0.67552	16
DLRG_RR	1	0.54180	17

bust metric particularly suited for imbalanced classification tasks, ensuring a fair assessment of performance across all classes. Thirty submissions in all were received, and each participant used a different method to identify misogyny and non-misogyny memes in multimodal contexts in Tamil and Malayalam. As per the Tamil findings displayed in Table 1, DLRG_RR obtained the highest rank with a macro F1 score of 0.83682. In order of precedence, CUET-NLP_Big_O came in second with 0.81716, byteSizedLLM with 0.80809, CUET-823 with 0.7812, and Dll5143 with 0.77591. CUET_Novice topped the Malayalam findings in Table 2 with an exceptional macro F1 score of 0.87631. With corresponding scores of 0.874833, 0.86658, 0.84927 and 0.84237, HerWILL, One_by_zero, Dll5143, and MNLP all shown strong performance.

The diverse methodologies employed by teams such as teamiic, byteSizedLLM, Team_Strikers, InnovationEngineers, Zero_knowledge, Code_Conquerors, HERWILL, Shraddha, CUET-NLP__Big_O, CUET-823 ,One_by_zero, CUET-NLP_MP and CUET_NetworkSociety highlighted the significance of multimodal approaches, combining textual and visual features for effective classification. Many teams leveraged pre-trained language models such as IndicBERT, XLM-RoBERTa, MuRIL, and multilingual BERT for textual feature extraction, often fine-tuned for Tamil and Malayalam. To capture visual contexts on the image side, models such as ResNet, Vision Transformer (ViT), Swin Transformer, and EfficientNet were frequently employed. In order to successfully integrate textual and visual embeddings, fusion techniques included dynamic attention mechanisms as well as early and late fusion where used.

The team Dll5143 improved the performance of the model by ensemble methods like majority voting and concatenation of multimodal embeddings, while teams such as buteSizedLLM and CUET-823 used novel approaches such as transliteration-enhanced datasets, back-translation, and data augmentation for both text and images, tackled issues with low-resource languages and imbalanced datasets. models with attention mechanisms, BiLSTM, and GRU captured contextual subtleties, and most of the teams employed dropout regularization and adam optimizer along with other hyperparameters. Teams such as Shraddha and LexiLogic used focal loss addressed overfitting and class imbalance.

The teams such as Magma, LinguAIsts and DLRG_RR used text based model like SVM, mBERT and BERT for classifiction task along with optimizing parameters. The team semanticCuetSync leveraged the LLaMa 3.2 , a large vision model for classification.

6 Conclusion and Future Work

In this second shared task, we aimed to address the issue of identifying misogyny memes in Tamil and Malayalam languages. The results demonstrated that multimodal fusion-based techniques yield better results in both the Tamil and Malayalam language dataset when compared with other techniques. Among the teams that submitted the results, most of them extracted textual and visual features separately using their appropriate models such as XLM-RoBERTa, T5, IndicBERT, MURIL for textual features and ResNet18, ViT, CNN, EfficientNetB0 for visual features. The features are then has been combined using fusion techniques and fed to the classifier model. Even though, multimodal models performed well for this dataset, we plan to explore the bias like data, contextual and algorithmic in data. Models can be trained to understand the local cultural differences and sensitivity in data by annotating the data set with detailed context information. Fine grained multimodal analysis is needed for detecting misogyny memes because of the subtle change in tone, context and nuances present in the image and text present in the memes which cannot be detected on the surface level analysis. Future research in these areas can improve the detection of misogyny memes for the safer online environment.

Acknowledgments

This work was conducted with the financial support from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2(Insight_2), supported in part of Science Foundation Ireland Centre for Research Training in Artificial Intelligence under Grant No. 18/CRT/6223.

References

Arthir, Pavithra J, G Manikandan, Lekhashree A4 Dhanyashree G, Bommineni Sahitya, Arivuchudar K, and Kalpana K. 2025. LinguAIsts@DravidianLangTech 2025: Misogyny Meme Detection using multimodel Approach. In *Proceedings of the Fifth Workshop on Speech,* *Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Valerio Basile, Cristina Bosco, Elisabetta Fersini, Debora Nozza, Viviana Patti, Francisco Manuel Rangel Pardo, Paolo Rosso, and Manuela Sanguinetti. 2019. SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 54–63, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Dola Chakraborty, Shamima Afroz Mithi, Jawad Hossain, and Mohammed Moshiul Hoque. 2025. One_by_zero@DravidianLangTech 2025: A Multimodal Approach for Misogyny Meme Detection in Malayalam Leveraging Visual and Textual Features. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Saranya Rajiakodi, Rahul Ponnusamy, Kathiravan Pannerselvam, Anand Kumar Madasamy, Ramachandran Rajalakshmi, Hariharan LekshmiAmmal, Anshid Kizhakkeparambil, Susminu S Kumar, Bhuvaneswari Sivagnanam, and Charmathi Rajkumar. 2024. Overview of Shared Task on Multitask Meme Classification - Unraveling Misogynistic and Trolls in Online Memes. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 139–144, St. Julian's, Malta. Association for Computational Linguistics.

Shraddha Chauhan and Abhinav Kumar. 2025. MNLP@DravidianLangTech 2025: Transformer-based Multimodal Framework for Misogyny Meme Detection. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the 37th International Conference on Machine Learning*, pages 1597–1607. PMLR. ISSN: 2640-3498.

Md. Sajid Alam Chowdhury, Mostak Mahmud Chowdhury, Anik Mahmud Shanto, Jidan Al Abrar, and Hasan Murad. 2025. Fired_from_NLP@DravidianLangTech 2025: A Multimodal Approach for Detecting Misogynistic Content in Tamil and Malayalam Memes. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Elisabetta Fersini, Debora Nozza, and Paolo Rosso. 2018. Overview of the Evalita 2018 Task on Automatic Misogyny Identification (AMI). In Tommaso Caselli, Nicole Novielli, and Viviana Patti, editors, *EVALITA Evaluation of NLP and Speech Tools for Italian : Proceedings of the Final Workshop 12-13 December 2018, Naples*, Collana dell'Associazione Italiana di Linguistica Computazionale, pages 59–66. Accademia University Press, Torino. Code: EVALITA Evaluation of NLP and Speech Tools for Italian : Proceedings of the Final Workshop 12-13 December 2018, Naples.

Francesca Gasparini, Giulia Rizzi, Aurora Saibene, and Elisabetta Fersini. 2022. Benchmark dataset of memes with text transcriptions for automatic detection of multi-modal misogynistic content. *Data in Brief*, 44:108526.

Md. Refaj Hossan, Nazmus Sakib, Md. Alam Miah Jawad Hossain Hoque, and Mohammed Moshiul. 2025. CUET-NLP_Big_O@DravidianLangTech 2025: A Multimodal Fusion-based Approach for Identifying Misogyny Memes. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Jianzhao Huang, Hongzhan Lin, Liu Ziyan, Ziyang Luo, Guang Chen, and Jing Ma. 2024. Towards low-resource harmful meme detection with LMM agents. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 2269–2293, Miami, Florida, USA. Association for Computational Linguistics.

Christos Koutlis, Manos Schinas, and Symeon Papadopoulos. 2023. MemeFier: Dual-stage Modality Fusion for Image Meme Classification. In *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, ICMR '23, pages 586–591, New York, NY, USA. Association for Computing Machinery.

Gitanjali Kumari, Kirtan Jain, and Asif Ekbal. 2024. M3Hop-CoT: Misogynous meme identification with multimodal multi-hop chain-of-thought. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22105–22138, Miami, Florida, USA. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint*. ArXiv:1907.11692 [cs].

Niranjan Kumar M, Pranav Gupta, Billodal Roy, and Souvik Bhattacharyya. 2025. LexiLogic@DravidianLangTech 2025: Detecting Misogynistic Memes and Abusive Tamil and Malayalam Text Targeting Women on Social Media. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Sidharth Mahesh, Sonith D, Gauthamraj Gauthamraj, Kavya G, Asha Hegde, and H Shashirekha. 2024. MUCS@LT-EDI-2024: Exploring joint representation for memes classification. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 282–287, St. Julian's, Malta. Association for Computational Linguistics.

Arpita Mallik, Ratnajit Dhar, Udoy Das, Momtazul Arefin Labib, Samia Rahman, and Hasan Murad. 2025. CUET-823@DravidianLangTech 2025: Shared Task on Multimodal Misogyny Meme Detection in Tamil Language. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Durga Prasad Manukonda and Rohith Gowtham Kodali. 2025. byteSizedLLM@DravidianLangTech 2025: Multimodal Misogyny Meme Detection in Low-Resource Dravidian Languages Using Transliteration-Aware XLM-RoBERTa, ResNet-50, and Attention-BiLSTM. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Md. Mohiuddin, Md Minhazul Kabir, Kawsar Ahmed, and Mohammedmoshiul Hoque. 2025. CUET-NLP_MP@DravidianLangTech 2025: A Transformer-Based Approach for Bridging Text and Vision in Misogyny Meme Detection in Dravidian Languages. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Sarbajeet Pattanaik, Ashok Yadav, and Vrijendra Singh. 2025. Dll5143@DravidianLangTech 2025: Majority Voting-Based Framework for Misogyny Meme Detection in Tamil and Malayalam. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Parth Patwa, Sathyanarayanan Ramamoorthy, Nethra Gunti, Shreyash Mishra, S Suryavardan, Aishwarya Reganti, Amitava Das, Tanmoy Chakraborty, Amit Sheth, Asif Ekbal, et al. 2022. Findings of memotion 2: Sentiment and emotion analysis of memes. In *Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR*.

Rahul Ponnusamy, Kathiravan Pannerselvam, Saranya R, Prasanna Kumar Kumaresan, Sajeetha Thavareesan, Bhuvaneswari S, Anshid K.a, Susminu S Kumar, Paul Buitelaar, and Bharathi Raja Chakravarthi. 2024. From laughter to inequality: Annotated dataset for misogyny detection in Tamil and Malayalam memes. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7480–7488, Torino, Italia. ELRA and ICCL.

Neelima Monjusha Preeti, , Trina Chakraborty, , Noor Mairukh Khan Arnob, , Saiyara Mahmud, , and Azmine Toushik and Wasi. 2025. HerWILL@DravidianLangTech 2025: Ensemble Approach for Misogyny Detection in Memes Using Pretrained Text and Vision Transformers. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8748–8763. PMLR. ISSN: 2640-3498.

Pathange Omkareshwara Rao, Harish Vijay V, Ippatapu Venkata Srichandra, Neethu Mohan, and Sachin Kumar S. 2025. Code_Conquerors@DravidianLangTech 2025: Multimodal Misogyny Detection in Dravidian Languages Using Vision Transformer and BERT. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Giulia Rizzi, Alessandro Astorino, Paolo Rosso, and Elisabetta Fersini. 2024. Unraveling disagreement constituents in hateful speech. In *European Conference on Information Retrieval*, pages 21–29. Springer.

Khadiza Sultana Sayma, Farjana Alam Tofa, Md Osama Dey, and Ashim. 2025. CUET_Novice@DravidianLangTech 2025: A Multimodal Transformer-Based Approach for Detecting Misogynistic Memes in Malayalam Language. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Kogilavani Shanmugavadivel, Malliga Subramanian, Mohamed Arsath H, Ramya K, and Ragav R. 2025a. TEAM_STRIKERS@DravidianLangTech2025: Misogyny Meme Detection in Tamil Using Multimodal Deep Learning. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Kogilavani Shanmugavadivel, Malliga Subramanian, Pooja Sree M, Palanimurugan V, and Roshini Priya K. 2025b. InnovationEngineers@DravidianLangTech 2025: Enhanced CNN Models for Detecting Misogyny in Tamil Memes Using Image and Text Classification. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Chhavi Sharma, Deepesh Bhageria, William Scott, Srinivas PYKL, Amitava Das, Tanmoy Chakraborty, Viswanath Pulabaigari, and Björn Gambäck. 2020. SemEval-2020 task 8: Memotion analysis- the visuo-lingual metaphor! In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 759–773, Barcelona (online). International Committee for Computational Linguistics.

Harshita Sharma, Simran, Vajratiya Vajrobol, and Nitisha Aggarwal. 2025. teamiic@DravidianLangTech 2025: Transformer-Based Multimodal Feature Fusion for Misogynistic Meme Detection in Low-Resource Dravidian Language. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Smriti Singh, Amritha Haridasan, and Raymond Mooney. 2023. “female astronaut: Because sandwiches won‘t make themselves up there”: Towards multimodal misogyny detection in memes. In *The 7th Workshop on Online Abuse and Harms (WOAH)*, pages 150–159, Toronto, Canada. Association for Computational Linguistics.

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Mihael Arcan, and Paul Buitelaar. 2020. Multimodal meme dataset (MultiOFF) for identifying offensive content in image and text. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 32–41, Marseille, France. European Language Resources Association (ELRA).

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Mihael Arcan, and Paul Buitelaar. 2023. TrollsWithOpinion: A taxonomy and dataset for predicting domain-specific opinion manipulation in troll memes. *Multimedia Tools and Applications*, 82(6):9137–9171.