

MUCS@LT-EDI-2024: Exploring Joint Representation for Memes Classification

Sidharth Mahesh^a, Sonith D^b, Gauthamraj^c,
Kavya G^d, Asha Hegde^e, H L Shashirekha^f

Department of Computer Science, Mangalore University, Mangalore, Karnataka, India
{^asidharthmaheshedu, ^bsonithksd, ^cgauthamrajdataspace}@gmail.com,
^dkavyamujk, ^ehegdekasha}@gmail.com, ^fhlsrekha@mangaloreuniversity.ac.in

Abstract

Misogynistic memes are a category of memes which contain disrespectful language targeting women on social media platforms. Hence, detecting such memes is necessary in order to maintain a healthy social media environment. To address the challenges of detecting misogynistic memes, "Multitask Meme classification - Unraveling Misogynistic and Trolls in Online Memes: LT-EDI@EACL 2024" shared task organized at European Chapter of the Association for Computational Linguistics (EACL) 2024, invites researchers to develop models to detect misogynistic memes in Tamil and Malayalam. The shared task has two sub-tasks and in this paper, we - team MUCS, describe the learning models submitted to Task 1 - Identification of Misogynistic Memes in Tamil and Malayalam. As memes represent multi-modal data of image and text, three models: i) Bidirectional Encoder Representations from Transformers (BERT)+Residual Network (ResNet)-50, ii) Multilingual Representations for Indian Languages (MuRIL)+ResNet-50, and iii) multilingual BERT (mBERT)+ResNet-50, are proposed based on joint representation of text and image, for detecting misogynistic memes in Tamil and Malayalam. Among the proposed models, mBERT+ResNet-50 and MuRIL+ ResNet-50 models obtained macro F1 scores of 0.73 and 0.87 for Tamil and Malayalam datasets respectively securing 1st rank for both the datasets in the shared task.

1 Introduction

Mememes, in the digital age, have become a common form of cultural expression, often shared widely across social media platforms and internet communities. These memes typically comprising of images/videos and text embedded on them, started with the idea of sharing humor (Suryawanshi et al., 2020). But these days, memes are often being misused to spread hateful, troll, and misogynistic content. Misogynistic memes are a category of memes

that propagate negative attitude towards women. These memes often promote dangerous or harmful pranks, challenges, or behaviors which leads to physical harm, injury, or legal consequences (Guest et al., 2021; Hegde et al., 2021). Hence, it is necessary to detect such content to protect users from getting harmed and also to maintain a safe and inclusive online environment.

Detecting misogynistic memes on social media is challenging due to the combination of text, image/video, and sometimes audio also, which exhibits a multi-modal nature. This problem becomes more challenging when the embedded text belongs to low-resource languages like Tamil, Malayalam etc., where lack of digital resources and computational tools is the common issue. "Multitask Meme Classification - Unraveling Misogynistic and Trolls in Online Memes: LT-EDI@EACL 2024" (Chakravarthi et al., 2024) shared task encourages the researchers to develop models to detect misogynistic and trolling content in Tamil and Malayalam memes. The shared task has two sub-tasks and in this paper, we - team MUCS, describe the learning models submitted to Task 1 - Identification of Misogynistic Memes in Tamil and Malayalam. As memes are made up of textual and visual components, they can be represented as multi-modal data of textual and visual features integrated into a single representation known as joint representation. Three models: i) BERT+ResNet-50, ii) MuRIL+ResNet-50, and iii) mBERT+ResNet-50, are proposed based on joint representation, for detecting misogynistic memes in Tamil and Malayalam.

The rest of the paper is arranged as follows: a review of related work is included in Section 2 and the methodology is discussed in Section 3. Experiments and results are described in Section 4 followed by concluding the paper with future work in Section 5.

2 Related work

Researchers have explored several models for detecting memes by representing the visual and textual components of memes as two uni-modal data as well as integrating visual and textual components into a single joint representation. Few of such relevant research works are described below: [Raha et al. \(2022\)](#) have explored uni-modal (Image-Grid, Image-Region, Text BERT, Text Robustly Optimized BERT Pre-training Approach (RoBERTa), Uni-modal fusions (Concat-BERT, Late Fusion), Multi-modal transformers (Multi-Modal BiTransformer (MMBT)-Grid, MMBT-Region, Vision-and-Language BERT (ViLBERT), Visual BERT) and pre-trained models (ViLBERT CC, Visual BERT COCO, ViLBERT HM, Visual BERT HM), for identifying misogynous memes in Conceptual Captions (CC), Common Objects in Context (COCO), Hateful Memes (HM) datasets. Among all the proposed models, the ViLBERT HM model outperformed all other models obtaining macro F1 score of 0.712 for HM dataset. [Muti et al. \(2022\)](#) proposed uni-modal and multi-modal approaches for identifying misogynistic memes in English dataset. The multi-modal system is implemented by fusing image and text embeddings through MMBT which is used to jointly fine-tune uni-modal pre-trained text and image encoders by projecting image embeddings to the text token space. Their proposed multi-modal system obtained a macro average F1 score of 0.727.

[Maheshwari and Nangi \(2022\)](#) experimented various Machine Learning (ML), Deep Learning (DL) and Transfer Learning (TL) based models for the identification of misogynous memes in English. Their ML models (Support Vector Machine (SVM), Naive Bayes (NB) and Logistic Regression (LR)) are trained with Term Frequency-Inverse Document Frequency (TF-IDF) of word unigrams (text representation) and pre-trained Visual Geometry Group-16 (VGG-16) (image representation), DL models (LSTM and Convolutional Neural Network (CNN)) are trained with GloVe embeddings (text representation) and VGG-16 (image representation) and TL models are trained with BERT variants (Concat BERT, Average BERT, and Gated BERT) (text representation) and Common World Knowledge (CWK) and Contrastive Language-Image Pre-training (CLIP) (image representation). The authors experimented all the models with uni-modal feature space, i.e., training the classifiers with only

text and only image features and also with joint learning i.e., training the classifiers with shared embedding layer for both text and image features. Among all their models, TL model with joint learning using Average BERT + CLIP achieved a macro F1 score of 0.671.

[Sean and Kanchana \(2022\)](#) presented multi-modal models, namely InceptionV3+BERT backbone as Model A, EfficientNetB7+BERT as Model B, CLIP Image+CLIP Text Backbone as Model C, SVM and an Ensemble model (Model A, Model B, SVM), for identifying misogynous memes in English. Among the proposed models, Ensemble model achieved a macro F1 score of 0.718. [Rao and Rao \(2022\)](#) experimented text-based (Bidirectional Long Short Term Memory (BiLSTM)+Glove embeddings, RoBERTa, Ernie-2.0), image-based (VGG-16, ResNet-50, ResNet-152, Vision Transformer), and multi-modal (VGG-16+BiLSTM, MMBT, VisualBERT, MMBT with tRoBERTa, and Average (Avg) Ensemble (RoBERTa and ResNet-152 models with soft voting)) models, for misogynous meme identification in English language. Among their proposed models, the Avg Ensemble model outperformed other models with a macro F1 score of 0.761. [Gu et al. \(2022\)](#) employed an ensemble of ML models (Multinomial NB (MNB) and Gradient Boosting classifiers trained with TF-IDF of word bigrams and unigrams respectively, and Random forest (RF) classifier trained with various image features (Hu moment invariants, Haralick textures, and image histograms), for the identification of misogynous memes. In addition, the ML models of the ensemble are also trained independently with the respective features as mentioned. Among all their models, RF classifier trained with various image features outperformed other models by achieving a macro F1 score of 0.665.

The above literature reveals that the joint representation of image and text exhibits promising performances for meme detection tasks. However, most of the meme detection tasks focus on English language giving less importance for low-resource languages like Tamil and Malayalam.

3 Methodology

The objective of this work is to identify misogynistic memes in the given Tamil and Malayalam datasets. This is achieved by proposing learning models based on the joint representation of image and text components in the given memes. The steps

Language	Image	Text	English Translation	Label
Malayalam		ഭാര്യ അയ്യപ്പിട്ട് ഒന്നും നടക്കുന്നില്ല എന്ന് ഇടക്കിടക്ക് സങ്കടം പറയുന്ന മുതലാളിയുടെ സുമിലേക്ക് കടന്നപ്പോൾ ഭാര്യയെ ഒറ്റക്കാലിൽ നിർത്തി കളിക്കുന്നത് കണ്ട വേലക്കാരീ* നിർത്തിയങ്ങു അടിക്കുവായിരുന്നല്ലേ...	When she entered the boss's room, who often complained that nothing was going on as a wife, the maid stopped and saw her playing on one leg.	Misogyny
		ഐഡിയ 4G ഹമ്പിൽ - മലയാളം വീടിന്റെ മുറ്റത്ത് 2G വിടിന്റെ ഹാളിൽ എന്റെ റൂമിൽ	Idea 4G Humpil - Malayalam In the yard of the house 2G in the hall of the house in my room	Not Misogyny
Tamil		~മാമിയാർ ശമൈക്കക തെറിയുമാ? പുതുമരുകൾ തെറിയുമാവാ അപ്പുപ പത്തവഴു കുടുതண்ணിയാ കായവഴു മേകി പക്കെട്ടെ കൊട്ടി അപ്പു കരണ്ടിയ വഴു അടവം	~ Does mother-in-law know how to cook? Do you know, newlyweds that the stove is heated, the water is heated, the water is dry, the bucket is poured, and the spoon is cleaned?	Misogyny
		നട എരിശാലും തിരുംപ തിരുംപ എழுந்து വര phoenix bird എങ്ക SILUKKU DOT COM முட்டை போட்டதுக்கே குண்டி வலக்குதுனுபுலம்புற நீ எங்க	Get up again and again with irritation phoenix bird where sila silukku dot com Where are you after laying the eggs and the pain in the stomach?	Not Misogyny

Table 1: Sample Malayalam and Tamil memes (image and text data) with corresponding labels

Tamil		
Label	Train set	Dev set
Misogyny	274	76
Not Misogyny	863	209
Malayalam		
Misogyny	256	64
Not Misogyny	384	96

Table 2: Class-wise distribution of memes in Tamil and Malayalam datasets

involved in the methodology are explained below:

3.1 Pre-processing

Pre-processing is a crucial step that cleans the data and prepares it for further processing. Usually, images will be of varying sizes as they will be collected from different sources and hence they are resized to a standard size. Further, images not in RGB format are converted to RGB format. Punctuation, digits, urls, and hashtags are considered as noise and hence are removed from the textual component. English stopwords (memes may also include English words), available at NLTK¹ library and Tamil stopwords from a GitHub² repository are utilized as references for removing English and Tamil stopwords from Tamil dataset respec-

tively and only English stopwords are removed from Malayalam datasets.

3.2 Construction of Learning Models

In DL, feature extraction and classifier construction go hand-in-hand. As memes contain image and embedded text, a joint representation of integrating image and text features is used in this work. The image and text encoders used to represent image and text respectively are described below:

- **Text Representation** - Transformer models have emerged as promising pre-trained models for extracting features from text due to their ability to capture intricate contextual relationships between words in the given input sequence. Their self-attention mechanisms enable a comprehensive understanding of word dependencies, allowing for the creation of context-rich embeddings that enhance the performance of many downstream natural language processing tasks. In this work, BERT³ (Devlin et al., 2018), MuRIL⁴ (Khanuja et al., 2021), and mBERT⁵ (Devlin et al., 2018), are used to represent text as three different models. BERT is pre-trained on a large amount of English text in a self-supervised fashion

¹<https://pythonspot.com/nltk-stop-words>

²<https://gist.github.com/arulrajnet/e82a5a331f78a5cc9b6d372df13a919c>

³<https://huggingface.co/bert-base-uncased>

⁴<https://huggingface.co/google/muril-base-cased>

⁵<https://huggingface.co/bert-base-multilingual-cased>

Language	Tamil					
Model	Dev set			Test set		
	Precision	Recall	F1 score	Precision	Recall	F1 score
BERT+ResNet-50	0.72	0.70	0.71	0.68	0.65	0.66
MuRIL+ResNet-50	0.77	0.72	0.74	0.75	0.64	0.66
mBERT+ResNet-50	0.75	0.72	0.73	0.77	0.72	0.73
Language	Malayalam					
BERT+ResNet-50	0.82	0.80	0.81	0.82	0.83	0.82
MuRIL+ResNet-50	0.86	0.80	0.81	0.90	0.87	0.87
mBERT+ResNet-50	0.77	0.75	0.76	0.85	0.84	0.84

Table 3: Performances of the proposed models for identifying misogynistic memes in Tamil and Malayalam datasets

using a Masked Language Modeling (MLM) objective whereas MuRIL and mBERT are multilingual pre-trained models which support Tamil and Malayalam languages. While MuRIL supports 17 Indian languages in their native and transliterated scripts, mBERT supports 104 languages in their native script. BERT is used as the given Tamil and Malayalam datasets contain English texts along with Tamil and Malayalam text in their native script.

- **Image Representation** - ResNet-50⁶ (He et al., 2016) - a CNN with 48 Convolution layers along with 1 Max Pool and 1 Average Pool layer and a fully connected layer, is a variant of ResNet which is pre-trained on ImageNet (Deng et al., 2009) dataset at a resolution of 224x224. ResNet-50 is used as image encoder to obtain the image features.

Dual-encoder architecture which is based on joint representation approach is used to concatenate image and text encoders and the joint encodings are passed through linear layers to build the classifier model for identifying misogynistic memes in Tamil and Malayalam.

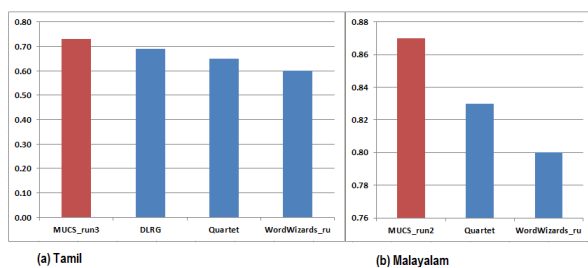


Figure 1: Comparison of macro F1 scores of the participating teams in the shared task

⁶<https://iq.opengenus.org/resnet50-architecture/>

4 Experiments and Results

Tamil and Malayalam memes datasets provided by the organizers of the shared task are labeled as 'Misogyny' and 'Not Misogyny' memes, for the task of binary classification (Chakravarthi et al., 2024). The sample memes with their corresponding labels and class-wise distribution of Tamil and Malayalam memes datasets are shown in Tables 1 and 2 respectively. Table 3 shows the performances of the proposed models. Among the proposed models, mBERT+ResNet-50 and MuRIL+ResNet-50 models obtained better macro F1 scores of 0.73 and 0.87 for Tamil and Malayalam datasets respectively, securing 1st rank for both the datasets in the shared task. Figure 1 gives a comparison of macro F1 scores of all the participating teams in the shared task.

4.1 Error Analysis

Few misclassified memes along with the actual and predicted labels obtained from mBERT+ResNet-50 and MuRIL+ResNet-50 for Tamil and Malayalam datasets respectively, are shown in Table 4. Misclassifications are due to the limitations of image and text encoders. Text encoders may fail to capture the domain specific meaning of the ambiguous words. Further, there are a few content words or phrases that are often used in the context of one polarity; however, the ground truth of the Test data with same words or phrases has a different polarity. For example, during training, the words or phrases 'quick', 'show', and 'in front of' are often used in the context of 'Misogyny' and the ground truth of this transcription is 'Not Misogyny'. From the image point of view, features that affect the identification of misogynistic memes include noise in the image, image quality, size of the training image dataset, and architecture of the image encoder.





Language	Tamil		Malayalam	
Meme				
Transcription	South Indian Aunties during சண்டை & குழாயடி சண்டை நீஎவன் கூட எல்லாம் தொடர்பு வச்சிருக்கனு எனக்கு தெரியும்டி தெரு	அவ என்னை ஏமாத்துனது கூட மன்னிச்சிருவேன் டா ஆனா? அவ புள்ளய விட்டு என்னை மாமாயன்னு கூப்பிட வச்சா பாரு ~ மறக்கவே முடியல dude	ബുസും പാവായുമുടുത്തു ഇള സിൻ കാണുമ്പോൾ	പെട്ടെന്ന് ദേഷ്യപ്പെടുന്നവർ, എത്ര വലിയ കലിപ്പ് കാണിച്ചാലും ആ കലിപ്പ് അവർ ഇഷ്ടപ്പെടുന്നവരുടെ മുന്നിൽ മാത്രമായിരിക്കും കാണിക്കുളള!
English Translation	South Indian Aunties during fight & Pipe fight I know that even you are connected to the street	Even if she cheated on me, I will forgive her. She left the room and called me Mamayan ~ never forget dude	Seeing Sin in a blouse and skirt	People who are quick to anger, no matter how much anger they show, that anger is only shown in front of those they love!
Actual Label	Misogyny	Not Misogyny	Misogyny	Not Misogyny
Predicted Label	Not Misogyny	Misogyny	Not Misogyny	Misogyny

Table 4: Few misclassified Tamil and Malayalam memes with actual and predicted labels

Added to this is the imbalance nature of the given datasets where both Tamil and Malayalam datasets contain less number of 'Misogyny' memes.

5 Conclusion and Future Work

This paper describes, three models: i) BERT+ResNet-50, ii) MuRIL+ResNet-50, and iii) mBERT+ResNet-50, based on joint representation of text and image features, for detecting misogynistic memes in Tamil and Malayalam datasets, submitted by our team - MUCS to "Multitask Meme classification - Unraveling Misogynistic and Trolls in Online Memes: LT-EDI@EACL 2024" shared task. Among the proposed models, mBERT+ResNet-50 and MuRIL+ResNet-50 models obtained macro F1 scores of 0.73 and 0.87 for Tamil and Malayalam datasets respectively, securing 1st rank for both the datasets in the shared task. More efficient joint representations that improve the performance of the learning models will be explored further.

References

Bharathi Raja Chakravarthi, Saranya Rajiakodi, Rahul Ponnusamy, Kathiravan Pannerselvam, Anand Kumar Madasamy, Ramachandran Rajalakshimi, Har-

iharan RamakrishnaIyer LekshmiAmmal, Anshid Kizhakkeparambil, Susminu S Kumar, Bhuvanewari 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 and Inclusion*, Malta. European Chapter of the Association for Computational Linguistics.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. *Imagenet: A large-scale hierarchical image database*. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. *BERT: pre-training of deep bidirectional transformers for language understanding*. In *CoRR*, volume abs/1810.04805.

Qin Gu, Nino Meisinger, and Anna-Katharina Dick. 2022. QiNiAn at SemEval-2022 Task 5: Multi-Modal Misogyny Detection and Classification. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 736–741.

Ella Guest, Bertie Vidgen, Alexandros Mittos, Nishanth Sastry, Gareth Tyson, and Helen Margetts. 2021. An Expert Annotated Dataset for the Detection of Online Misogyny. In *Proceedings of the 16th Conference of the European Chapter of the Association for Com-*

- putational Linguistics: Main Volume*, pages 1336–1350.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Asha Hegde, Mudoor Devadas Anusha, Sharal Coelho, and Hosahalli Lakshmaiah Shashirekha. 2021. Mum at comma@ icon: Multilingual gender biased and communal language identification using supervised learning approaches. In *Proceedings of the 18th International Conference on Natural Language Processing: Shared Task on Multilingual Gender Biased and Communal Language Identification*, pages 64–69.
- Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, Shruti Gupta, Subhash Chandra Bose Gali, Vish Subramanian, and Partha Talukdar. 2021. [Muril: Multilingual representations for indian languages](#).
- Paridhi Maheshwari and Sharmila Reddy Nangi. 2022. TeamOtter at SemEval-2022 Task 5: Detecting Misogynistic Content in Multimodal Memes. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 642–647.
- Arianna Muti, Katerina Korre, and Alberto Barrón-Cedeño. 2022. UniBO at SemEval-2022 Task 5: A Multimodal bi-Transformer Approach to the Binary and Fine-grained Identification of Misogyny in Memes. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 663–672.
- Tathagata Raha, Sagar Joshi, and Vasudeva Varma. 2022. IITH at SemEval-2022 Task 5: A Comparative Study of Deep Learning Models for Identifying Misogynous Memes. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 673–678.
- Ailneni Rakshitha Rao and Arjun Rao. 2022. ASRtrans at SemEval-2022 Task 5: Transformer-based Models for Meme Classification. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 597–604.
- Benhur Sean and Sivanraju Kanchana. 2022. Transformers at SemEval-2022 Task 5: A Feature Extraction based Approach for Misogynous Meme Detection. In *Proceedings of the 16th International Workshop on Semantic Evaluation*, pages 550–554.
- Shardul Suryawanshi, Bharathi Raja Chakravarthi, Michael 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.