Behind the Laughter: Uncovering Gender Bias in Code-Mixed Bangla Memes

Jannatul Ferdusi[1], **Labanya Saha**[1], **Paria Chowdhury**[1], **Jawad Hossain**[2], and **Noor Mairukh Khan Arnob**[1*]
[1]University of Asia Pacific, Dhaka-1205, Bangladesh
[2]University at Albany - State University of New York, USA

Abstract

Bangla memes are widely used on social media to express humor and social commentary, yet computational analysis of gender bias in Bangla memes remains largely unexplored. In this work, we present a multimodal framework for detecting gender bias in Bangla memes by jointly analyzing textual and visual content. We construct a dataset of 6,846 Bangla and Banglish code-mixed memes annotated into three categories: male-biased, female-biased, and neutral. For textual representation, we use BanglishBERT, while visual features are extracted using ConvNeXt, and the two modalities are fused for final classification. Our best-performing model, ConvNeXt + BanglishBERT, achieves accuracy of 0.67 and an F1-score of 0.63, outperforming several multimodal baselines. The results demonstrate the effectiveness of multimodal learning for understanding culturally nuanced and code-mixed meme content in low-resource languages. Code and data available at this link.

1 Introduction

Memes have emerged as a powerful form of digital communication on social media, combining images and text to convey humor, opinions, and social commentary. While memes often serve as entertainment, studies show that they are also frequently used to express offensive, hateful, or biased narratives through the interaction of visual and textual cues (Chen and Pan, 2022). In Bangladeshi online communities, Bangla memes have become increasingly popular and often reflect social attitudes and cultural norms. Recent discussions on online harms highlight the risks of tech-facilitated gender-based violence (TFGBV), where digital media such as memes and images can be used to propagate gendered harassment and abuse, emphasizing the need for automated detection systems (Mobashwira, 2026). Analyzing such content can therefore provide insights into how gender roles and social perceptions are represented in digital culture.

Prior research has demonstrated that multimodal approaches significantly improve meme understanding by jointly analyzing visual and textual information. Early studies showed that combining these modalities improves the detection of harmful or offensive meme content (Oriol Sàbat, 2019). Similarly, visual–linguistic pre-training and multimodal architectures have been successfully applied to tasks such as offensive meme detection, hate speech identification, and sentiment analysis (Hakimov et al., 2024). In the Bangla context, existing work has mainly focused on sentiment analysis (Ahammad et al., 2024) and misogyny detection (Mia et al., 2025), highlighting both the importance and challenges of analyzing memes in low-resource languages.

Despite these advances, the computational analysis of *gender bias direction* in Bangla memes remains largely unexplored. Most existing systems focus on detecting whether a meme is hateful or offensive but do not explicitly identify whether the bias targets males, females, or neither (Mia et al., 2025). Furthermore, humor, bias, and sentiment in memes are often shaped by cultural context, making models trained on high-resource languages difficult to apply directly to Bangla meme data (Xie et al., 2023). These challenges highlight the need for dedicated datasets and models for Bangla meme analysis.

To address these limitations, we propose a multimodal framework for detecting gender bias in Bangla memes by jointly analyzing textual and visual information. We construct a dataset of Bangla, Banglish, and code-mixed memes collected from social media and annotate them into three categories: *male-biased*, *female-biased*, and *neutral*. Our model combines BanglishBERT for textual representation with ConvNeXt for visual feature

*Correspondence: arnob@uap-bd.edu

Proceedings of the Sixth Workshop on Language Technology for Equality, Diversity, Inclusion, pages 1–9

extraction through a multimodal fusion mechanism. The system is evaluated using standard metrics including accuracy, precision, recall, and F1-score.

The main contributions of this work are summarized as follows:

- We develop a multimodal framework for detecting gender bias in Bangla-Banglish memes by jointly analyzing textual and visual information.
- We construct a dataset of Bangla memes annotated for gender bias direction, consisting of 6,846 memes with additional metadata.
- We provide statistical analysis and insights into gender bias patterns in Bangla meme culture on social media.

2 Related Work

Recent years have witnessed growing interest in analyzing harmful and biased content in memes using multimodal learning techniques. Memes typically combine images and short textual captions, making them inherently multimodal and challenging for traditional text-only or vision-only models. As a result, many studies have explored the use of computer vision and natural language processing methods to jointly analyze both modalities for detecting offensive, hateful, or biased content in memes.

One line of work focuses on detecting hate speech and offensive content in memes. For example, (Karim et al., 2022) investigate hate speech detection in Bengali memes using a multimodal deep learning framework that combines textual features extracted from NLP models with visual features obtained from convolutional neural networks (CNNs). Their study demonstrates that integrating visual and textual signals improves the detection of harmful meme content compared to unimodal approaches.

Several studies have also explored multimodal architectures for sentiment analysis in memes. For instance, (Faria et al., 2025) introduce *Sentiment-Former*, a transformer-based multimodal framework designed to analyze sentiment in Bangla memes. Their approach uses transformer models to extract semantic representations from text and CNN-based models to capture visual information from meme images. These representations are fused through a multimodal architecture to improve sentiment prediction performance.

Another closely related line of research focuses on identifying offensive or hateful meme content in Bangla. The work of (Nahin et al., 2024) presents a dataset and a deep learning framework for detecting hateful memes in Bengali. Their approach also uses multimodal representations combining visual and textual information to identify harmful content. Similarly, (Mia et al., 2025) introduce BANMIME, a dataset and benchmark for misogyny detection in Bangla memes, highlighting the challenges of identifying gender-targeted harassment in multimodal social media content.

Beyond Bangla, research on meme analysis has also expanded to other low-resource languages. For example, (Ponnusamy et al., 2024) present a multilingual dataset for detecting misogyny in memes across South Indian languages such as Tamil and Malayalam. Their work demonstrates that memes often encode culturally specific forms of humor and bias that require language-specific datasets and models.

More broadly, multimodal meme understanding has been widely studied in the NLP community. Early benchmark efforts such as the Hateful Memes dataset (Hossain et al., 2024) introduced large-scale multimodal datasets for detecting hateful content in Bangla memes. Subsequent research has explored various multimodal architectures, including vision–language transformers and cross-modal attention mechanisms, to better capture the interaction between images and text in meme content (Ahsan et al., 2024). These studies highlight the importance of jointly modeling textual and visual cues when analyzing social media memes.

Despite these advances, existing work primarily focuses on detecting whether a meme is hateful, offensive, or misogynistic. In contrast, relatively little research has examined the *direction of gender bias* in memes, particularly in low-resource languages such as Bangla. Most prior work treats the task as a binary classification problem (e.g., hateful vs. non-hateful), without distinguishing whether the bias targets men, women, or neither.

In contrast to previous studies, our work specifically focuses on detecting the *direction of gender bias* in Bangla memes by categorizing memes into three classes: male-biased, female-biased, and neutral. We also introduce a curated dataset of Bangla and Banglish code-mixed memes collected from social media and annotated for gender bias

direction. Furthermore, we propose a multimodal framework that integrates BanglishBERT for textual representation learning and ConvNeXt for visual feature extraction. By jointly analyzing visual and textual signals, our work aims to provide deeper insights into how gender bias manifests in Bangla meme culture and contributes a new benchmark for studying gender bias in low-resource multimodal content.

3 Dataset and Task

For this research, we developed a custom Bangla meme dataset to study gender bias in social media memes. Memes were collected from publicly available sources such as Facebook pages, public groups, and personal public profiles. Each meme consists of an image paired with an associated caption written in Bangla, Banglish, or code-mixed text. To balance reproducibility with responsible data sharing, the dataset is distributed under controlled access. Interested researchers may request access via this form.

3.1 Data Collection

The dataset was manually collected over a period of seven months and twenty-five days, from March 25, 2025 to November 19, 2025. During the collection process, we ensured that all memes were publicly available and that duplicate entries were removed. Memes were retrieved using keyword-based search strategies related to gender, memes, humor, and cultural expressions commonly used in Bangla-speaking online communities.

In addition to meme content, we also collected metadata about the uploader when available. This includes whether the meme was posted by a male user, female user, or a public Facebook page. This metadata enables further analysis of how gender bias may vary depending on the type of content creator.

The final dataset contains 6,846 memes in total, including 1,935 memes biased against males, 1,470 memes biased against females, and 3,441 neutral memes. The label distribution is shown in Table 1. The dataset and code used in this study will be released upon acceptance to facilitate further research.

3.2 Annotation Process

Each meme was manually annotated to identify the direction of gender bias present in the content.

Table 1: Distribution of labels in our dataset, indicating that more memes are biased against males.

Gender Bias Category	No. of samples
Male-biased (MaB)	1,935
Female-biased (FeB)	1,470
Neutral (Neu)	3,441
Total	**6,846**

The annotation process considered both the textual caption and the visual information contained in the meme image.

Two independent annotators with native proficiency in Bangla reviewed each meme and assigned one of three labels based on predefined annotation guidelines. In cases where disagreements occurred, a third annotator reviewed the meme and resolved the conflict through discussion. To measure annotation reliability, we computed Cohens Kappa coefficient between the two primary annotators, obtaining a score of $\kappa = 0.81$, which indicates strong agreement. This multi-stage annotation process helped ensure the reliability and consistency of the labels.

3.3 Label Definition

Each meme in the dataset was categorized into one of the following three classes:

- **Male-biased (MaB)**: Memes that contain negative stereotypes, ridicule, criticism, or derogatory humor targeting men.
- **Female-biased (FeB)**: Memes that contain stereotypes, mockery, or harmful narratives directed toward women.
- **Neutral (Neu)**: Memes that do not target any particular gender and contain general humor or commentary unrelated to gender bias.

3.4 Poster Category Analysis

To further analyze how gender bias appears across different types of content creators, we categorized memes based on the uploader type and the gender targeted in the meme. Table 2 presents this distribution.

In Table 2, the arrow symbol ($\rightarrow$) represents the relationship between the *poster type* and the *gender targeted in the meme*. Specifically, the left side of the arrow indicates the uploader of the meme (Male user, Female user, or Facebook Page), while

Table 2: Distribution of post categories by occurrence and percentage.

Post Category	Count	%
Male → Male	631	9.22
Male → Female	347	5.07
Male → Neutral memes	504	7.36
Female → Male	170	2.48
Female → Female	376	5.49
Female → Neutral memes	281	4.10
Page → Male	1134	16.56
Page → Female	748	10.93
Page → Neutral memes	2655	38.78

the right side indicates the gender group targeted by the meme (Male, Female, or Neutral). For example, *Male → Female* refers to memes posted by male users that contain bias targeting females.

The table reveals several patterns in Bangla meme culture. Facebook pages account for the largest share of memes overall. In particular, pages posting neutral memes constitute the largest category (38.78%), followed by pages posting memes targeting males (16.56%) and females (10.93%). This suggests that institutional or semi-anonymous pages play a significant role in shaping meme discourse.

Among individual users, male users contribute a higher volume of memes across most categories compared to female users. Male users post more antagonistic memes targeting both males (9.22%) and females (5.07%), as well as neutral content (7.36%). Female users show lower overall participation, with neutral memes (4.10%) slightly exceeding antagonistic content. Interestingly, both male and female users post more memes targeting their own gender than the opposite gender, suggesting the presence of intra-gender critique alongside inter-gender bias.

3.5 Example Memes

Table 3 presents example meme captions and their corresponding gender bias labels from the dataset, illustrating the diversity of linguistic expressions found in Bangla meme culture.

Figure 1 shows representative meme images from the dataset, demonstrating the diversity of visual styles and cultural references present in Bangla memes.

Table 3: Sample captions and gender bias labels from our Meme Dataset, showing varied cultural context.

SL No.	Text (Bangla Caption)	Bias
1	বুঝি না আমার বয়স ১৮ নাকি ৮০	Female
2	গরিব হতে পারি ঠিকই কিন্তু অরিজিনাল	Neutral
3	ভাই বেডি নিয়া ঝগড়া কইরেন না	Female
4	ক্রাশকে বললাম তোমার মন চুরি করতে	Male
5	তেলের যা দাম, নিজের চরকায় থুথু দিন	Neutral
6	নারী দিবসে কেউ উইশ করল না। নিজেকে	Female
7	ছেলেদের সাথে বন্ধুত্ব করি না কারণ দুইদিন গেলেই বলবে	Female

Figure 1: Sample meme images from our dataset, showcasing the visual diversity accompanied by cultural relevance.

4 Methodology

The proposed system adopts a multimodal architecture designed to analyze both textual and visual information present in Bangla memes. The framework consists of two parallel branches that independently process meme captions and meme images. Each branch extracts high-level feature representations using specialized encoders. These representations are then fused to form a unified multimodal embedding, which is subsequently used for gender bias classification. Figure 2 illustrates the overall architecture of the proposed system.

4.1 Pre-processing

Before feeding the data into the model, both image and text inputs are preprocessed to ensure con-

sistent representation and improved model performance.

Image preprocessing: Each meme image was first converted to RGB format and resized to 224×224, which is the required input size for the ConvNeXt architecture. The images were then normalized using the standard ImageNet mean and standard deviation. To improve model generalization across diverse meme formats, light data augmentation techniques were applied, including random horizontal flipping and mild brightness and contrast adjustments. These augmentations help the model handle variations in meme styles, fonts, and backgrounds.

Text preprocessing: The meme captions were cleaned by removing unnecessary symbols, repeated characters, and textual noise commonly found in social media content. After cleaning, the captions were tokenized using the BanglishBERT tokenizer, which performs subword tokenization and automatically applies padding and truncation to a fixed sequence length.

After preprocessing, both textual and visual inputs are converted into feature embeddings that are later combined through a multimodal fusion mechanism.

4.2 Textual Representation

For textual feature extraction, we experimented with two models: BanglishBERT (Bhattacharjee et al., 2022) and Sentence-BERT (Reimers and Gurevych, 2019). From the Sentence Transformer family, we selected the `paraphrase-multilingual-MiniLM-L12-v2` model due to its strong multilingual capabilities.

BanglishBERT was chosen as the primary textual encoder because it is specifically designed for Bangla and English bilingual text. Banglish, a code-mixed form of Bangla and English, is widely used in social media memes and often includes phonetic spellings, transliterations, and mixed-language expressions. BanglishBERT is therefore well-suited for capturing the linguistic characteristics of such data.

The textual processing pipeline begins with a text normalization step (Hasan et al., 2020), followed by tokenization. The tokens are then mapped into embedding vectors containing word, positional, and token-type representations. These embeddings are passed through multiple ELECTRA-based Transformer layers, which capture contextual relationships between words using self-attention mechanisms. The resulting contextual embeddings represent the semantic meaning of the meme captions and are used as input for downstream classification.

4.3 Visual Representation

For visual feature extraction, we use ConvNeXt (Liu et al., 2022), a modern convolutional neural network architecture that combines the efficiency of CNNs with design principles inspired by vision transformers.

The preprocessed meme images are passed through the ConvNeXt encoder, which processes images through four hierarchical stages. The first stage captures low-level visual features such as edges, colors, and simple patterns. The intermediate stages extract mid-level features such as textures, shapes, and structural patterns present in meme images. Finally, the last stage captures high-level semantic representations that describe objects, visual context, and scene-level information. These visual embeddings provide meaningful representations of the meme images for downstream classification.

4.4 Multimodal Fusion

To combine the information from both modalities, the feature embeddings obtained from BanglishBERT and ConvNeXt are concatenated to form a unified multimodal representation. This fused embedding captures both the semantic meaning of the textual captions and the contextual visual information contained in the meme image.

The combined feature vector is passed through a set of fully connected layers, followed by a softmax classifier that predicts the gender bias category of the meme (male-biased, female-biased, or neutral).

To evaluate the effectiveness of different model combinations, we experimented with six multimodal architectures formed by pairing different textual encoders (BanglishBERT and Sentence Transformer) with visual encoders (ConvNeXt, ResNet50, and ViT). Among these configurations, the combination of BanglishBERT and ConvNeXt achieved the best performance. All models were trained for 30 epochs, after which the validation accuracy stabilized.

5 Experimental Setup

All experiments were implemented using the PyTorch framework. The dataset was divided into

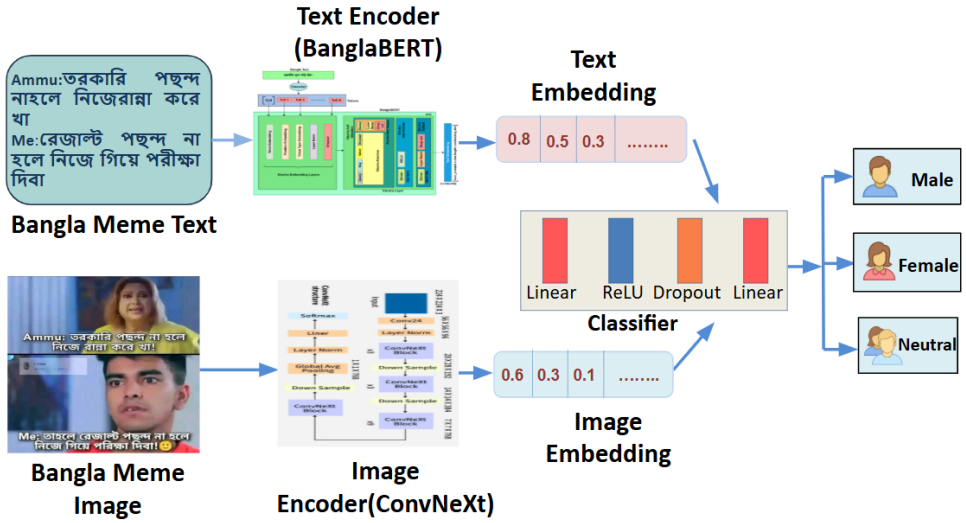

Figure 2: Block diagram of the proposed multimodal meme classification framework showing the text encoder, image encoder, and classifier module.

training (80%), validation (10%), and test (10%) sets using stratified sampling to maintain the original class distribution across the splits. To mitigate class imbalance during training, a weighted random sampler was applied to ensure balanced sampling of meme categories.

The multimodal architecture combines textual features extracted from BanglishBERT and visual features extracted from ConvNeXt-Base. BanglishBERT produces a textual embedding of dimension 768, while ConvNeXt-Base generates visual feature vectors of dimension 1,024. These embeddings are concatenated to form a unified multimodal representation, which is then passed through a two-layer fully connected classifier. The classifier uses a hidden layer of size 512 with ReLU activation and a dropout rate of 0.2 to reduce overfitting.

For textual inputs, meme captions were tokenized using the BanglishBERT tokenizer with a maximum sequence length of 64 tokens. Image inputs were resized to 224×224 pixels and normalized using the standard ImageNet mean and standard deviation.

The model was optimized using the AdamW optimizer with a learning rate of 2×10^{-5}. Training was conducted for 30 epochs with a batch size of 16. The best model checkpoint was selected based on validation accuracy and subsequently evaluated on the held-out test set.

Table 4: Performance comparison of multimodal models. Our proposed **ConvNeXt + BanglishBERT** achieves the best results across all evaluation metrics.

Model	Acc	Pr	Re	F1
ConvNeXt + Sentence-Transformer	0.56	0.53	0.48	0.48
ResNet50 + Sentence-Transformer	0.55	0.53	0.44	0.43
ResNet50 + BanglishBERT	0.65	0.63	0.60	0.61
ViT + BanglishBERT	0.66	0.65	0.60	0.61
ViT + Sentence-Transformer	0.57	0.56	0.48	0.49
ConvNeXt + BanglishBERT	**0.67**	**0.66**	**0.61**	**0.63**

6 Result Analysis

Table 4 presents the performance comparison of six multimodal architectures evaluated using accuracy, precision, recall, and F1-score. The models combine different visual encoders (ConvNeXt, ResNet50, and ViT) with textual encoders (BanglishBERT and Sentence Transformer).

Among the evaluated models, **ConvNeXt + BanglishBERT** achieved the best overall performance with an accuracy of **67%** and an F1-score of **0.63**. The model also maintained relatively strong precision (0.66) and recall (0.61), indicating balanced classification performance across classes. During training, the model reached a training accuracy of 92% with a training loss of 0.074 and a validation loss of 1.5552.

The second-best model was **ViT + BanglishBERT**, which achieved an accuracy of 66%. Models that utilized BanglishBERT consistently outperformed those using Sentence Transformer as the textual encoder. This suggests that BanglishBERT is more effective for handling Bangla and

Table 5: Training time (minutes) for each multimodal model. Sentence-Transformer based models generally train faster than BanglishBERT-based models.

Model	Time (min)
ConvNeXt + BanglishBERT	43
ConvNeXt + Sentence-Transformer	33
ResNet50 + Sentence-Transformer	39
ResNet50 + BanglishBERT	45
ViT + BanglishBERT	57
ViT + Sentence-Transformer	41

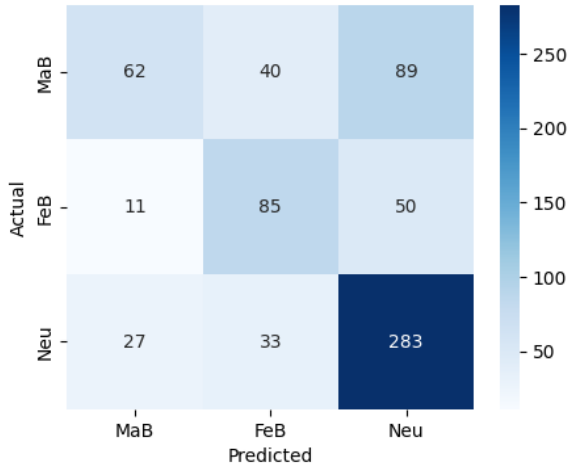

Figure 3: Confusion matrix of the proposed model, illustrating classification performance across gender bias categories.

code-mixed Banglish text commonly found in social media memes.

ResNet50-based models achieved moderate performance, with the combination of ResNet50 and BanglishBERT reaching an accuracy of 65%. In contrast, models paired with Sentence Transformer generally produced lower results across all visual encoders.

Table 5 compares the training time required for each multimodal model. Models using Sentence Transformer generally trained faster than those using BanglishBERT. However, despite the higher computational cost, BanglishBERT-based models consistently achieved better classification performance. Among all configurations, **ConvNeXt + BanglishBERT** achieved the best balance between performance and training efficiency, requiring 43 minutes for training. The **ViT + BanglishBERT** model required the longest training time (57 minutes), likely due to the higher computational complexity of the ViT architecture.

To further analyze the model's behavior, we examine the confusion matrix shown in Figure 3. The confusion matrix illustrates the distribution of correct and incorrect predictions across the three gender bias categories. Each row represents the true class, while each column represents the predicted class.

The results indicate that the model performs best on the **neutral class**, correctly identifying 283 samples with relatively few misclassifications. This suggests that neutral memes contain more consistent linguistic and visual patterns, making them easier for the model to recognize.

For the **male-biased class**, the model correctly classified 62 samples but misclassified several instances as female-biased or neutral. Similarly, for the **female-biased class**, the model correctly predicted 85 samples but also confused some instances with the neutral class. These misclassifications may occur because memes targeting different genders often share similar visual formats or sarcastic textual expressions.

Overall, the confusion matrix reveals that while the model performs well in detecting neutral memes, distinguishing between male-biased and female-biased memes remains more challenging. This highlights the complexity of detecting subtle gender bias in meme content, where humor, sarcasm, and cultural context often overlap across categories.

7 Acknowledgement

This research is funded by the Institute for Research, Innovation, and Development (IRID) at the University of Asia Pacific (UAP). We would also like to thank the Research and Innovation Lab, Department of Computer Science and Engineering, University of Asia Pacific for providing ample hardware support for conducting this research work.

8 Conclusion

In this work, we presented a multimodal framework for detecting gender bias in Bangla memes by jointly analyzing textual and visual content. We introduced a dataset of Bangla and Banglish code-mixed memes labeled as male-biased, female-biased, or neutral. Our approach combines BanglishBERT for textual representation and ConvNeXt for visual feature extraction to capture both linguistic and visual cues in meme content. Experimental results show that the proposed multimodal model outperforms several alternative architectures, highlighting the effectiveness of multimodal learning for analyzing culturally nuanced social media data in low-resource languages. In future work, we aim to explore larger multimodal models and cross-lingual transfer methods to further improve gender bias detection in meme content.

Limitations

Although the model achieves competitive performance but still struggles with subtle gender bias expressed through sarcasm. The lack of a large and high-quality Bangla meme dataset was a challenging issue, which made it difficult for the models to fully understand cultural expressions, sarcasm, and mixed language patterns. Since memes often contain Banglish and heavy code-mixing, many text exerpts were noisy or stylized, and this reduced the performance of text encoders. Some memes contained very small or distorted text that even struggled to detect correctly. Visual elements like low resolution, heavy filters, and complex backgrounds also affected feature extraction. The model sometimes got confused by sarcastic memes where the humor depends on real-world context. Class imbalance was another limitation because some categories had fewer samples, leading to biased predictions.

Ethics Statement

All data used in this study were collected exclusively from publicly accessible Facebook posts, pages, and groups with privacy settings set to public. No private, restricted, or access-controlled content was used.

The dataset contains only meme images and their corresponding annotation labels. No personally identifiable information (PII), including names, profile identifiers, contact details, or metadata that could directly identify individuals, was collected or retained. Although the perceived gender of content uploaders was recorded for aggregate statistical analysis, this information was not linked to any identifiable user data.

Given the sensitive nature of gender bias and the potential for misuse of social media content, the dataset is not publicly released. Instead, it is shared under controlled access. Researchers may obtain the dataset by submitting a request form and agreeing to a data use policy that restricts usage to non-commercial academic research, prohibits attempts to identify or contact individuals, and forbids redistribution.

All reasonable steps were taken to minimize potential harm, protect user privacy, and ensure responsible use of publicly available data. This work adheres to established ethical guidelines for social media research and data handling.

References

Tanzin Ahammad, Shawly Ahsan, Jawad Hossain, and Mohammed Moshiul Hoque. 2024. M-sam: Multimodal sentiment analysis exploiting textual and visual features of social media memes. In *International Conference on Pattern Recognition*, pages 134–150. Springer.

Shawly Ahsan, Eftekhar Hossain, Omar Sharif, Avishek Das, Mohammed Moshiul Hoque, and M Dewan. 2024. A multimodal framework to detect target aware aggression in memes. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2487–2500.

Abhik Bhattacharjee, Tahmid Hasan, Kazi Mubasshir, Md. Saiful Islam, Wasi Ahmad Uddin, Anindya Iqbal, M. Sohel Rahman, and Rifat Shahriyar. 2022. Banglabert: Lagnuage model pretraining and benchmarks for low-resource language understanding evaluation in bangla. In *Findings of the North American Chapter of the Association for Computational Linguistics: NAACL 2022*.

Yuyang Chen and Feng Pan. 2022. Multimodal detection of hateful memes by applying a vision-language pre-training model. *Plos one*, 17(9):e0274300.

Fatema Tuj Johora Faria, Laith H. Baniata, Mohammad H. Baniata, Mohannad A. Khair, Ahmed Ibrahim Bani Ata, Chayut Bunterngchit, and Sangwoo Kang. 2025. Sentimentformer: A transformer-based multimodal fusion framework for enhanced sentiment analysis of memes in under-resourced bangla language. *Electronics*, 14(4):799.

Sherzod Hakimov, Gullal S Cheema, and Ralph Ewerth. 2024. Processing multimodal information: Challenges and solutions for multimodal sentiment analysis and hate speech detection. In *Event Analytics across Languages and Communities*, pages 71–94. Springer.

Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, Masum Hasan, Madhusudan Basak, M. Sohel Rahman, and Rifat Shahriyar. 2020. Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for Bengali-English machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2612–2623, Online. Association for Computational Linguistics.

Eftekhar Hossain, Omar Sharif, Mohammed Moshiul Hoque, and Sarah Masud Preum. 2024. Align before attend: Aligning visual and textual features for multimodal hateful content detection. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 162–174.

Md Rezaul Karim, Sumon Kanti Dey, Tanhim Islam, Md Shajalal, and Bharathi Raja Chakravarthi. 2022. Multimodal hate speech detection from bengali memes and texts. In *International Conference on Speech and Language Technologies for Low-resource Languages*, pages 293–308. Springer.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. 2022. A convnet for the 2020s. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Md Ayon Mia, Akm Moshiur Rahman Mazumder, Khadiza Sultana Sayma, Md Fahim, Md Tahmid Hasan Fuad, Muhammad Ibrahim Khan, and Akmmahbubur Rahman. 2025. BANMIME : Misogyny detection with metaphor explanation on Bangla memes. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 17813–17839, Suzhou, China. Association for Computational Linguistics.

Jeba Mobashwira. 2026. Threats of tech-facilitated gender-based violence. The Daily Star, Law & Our Rights. Accessed: 11 March 2026.

Abrar Shadman Mohammad Nahin, Isfara Islam Roza, Tasnuva Tamanna Nishat, Afia Sumya, Hanif Bhuiyan, and Md Moinul Hoque. 2024. Bengali hateful memes detection: A comprehensive dataset and deep learning approach. In *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*, pages 01–06. IEEE.

Benet Oriol Sàbat. 2019. Multimodal hate speech detection in memes. B.S. thesis, Universitat Politècnica de Catalunya.

Rahul Ponnusamy, Kathiravan Pannerselvam, R Saranya, Prasanna Kumar Kumaresan, Sajeetha Thavareesan, S Bhuvaneswari, Anshid Ka, 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.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Heng Xie, Jizhou Cui, Yuhang Cao, Junjie Chen, Jianhua Tao, Cunhang Fan, Xuefei Liu, Zhengqi Wen, Heng Lu, Yuguang Yang, and 1 others. 2023. Multimodal cross-lingual features and weight fusion for cross-cultural humor detection. In *Proceedings of the 4th on Multimodal Sentiment Analysis Challenge and Workshop: Mimicked Emotions, Humour and Personalisation*, pages 51–57.