hate-alert@DravidianLangTech-2023: Multimodal Abusive Language Detection and Sentiment Analysis in Dravidian Languages

Shubhankar Barman
BITS Pilani, India
contact.shubhankarbarman@gmail.com

Mithun Das
IIT Kharagpur, India
mithundas@iitkgp.ac.in

Abstract

The use of abusive language on social media platforms is a prevalent issue that requires effective detection. Researchers actively engage in abusive language detection and sentiment analysis on social media platforms. However, most of the studies are in English. Hence, there is a need to develop models for low-resource languages. Further, the multimodal content in social media platforms is expanding rapidly. Our research aims to address this gap by developing a multimodal abusive language detection and performing sentiment analysis for Tamil and Malayalam, two under-resourced languages, based on the shared task "Multimodal Abusive Language Detection and Sentiment Analysis in Dravidian Languages: DravidianLangTech@RANLP 2023". In our study, we conduct extensive experiments utilizing multiple deep-learning models to detect abusive language in Tamil and perform sentiment analysis in Tamil and Malayalam. For feature extraction, we use the mBERT transformer-based model for texts, the ViT model for images and MFCC for audio. In the abusive language detection task, we achieved a weighted average F1 score of **0.5786**, securing the **first** rank in this task. For sentiment analysis, we achieved a weighted average F1 score of **0.357** for Tamil and **0.233** for Malayalam, ranking **first** in this task.

1 Introduction

Social media platforms have been expanding rapidly with a variety of content in different languages. On social media, users express their opinions with a few limitations, the majority of social media platforms allow users to share and express their thoughts and they aim to gather user comments and posts to offer a personalized feed. However, they are also used for negative activities, such as spreading rumors and bullying people with abusive words. Abuse of language has received a lot of attention as social media platforms have grown in popularity (Das et al., 2020; Banerjee et al., 2021; Das et al., 2021b). When someone uses language that is hurtful, disrespectful, or disparaging towards another person or group because of traits like race, ethnicity, gender, religion, sexual orientation, or other personal qualities, that language is considered abusive(Waseem et al., 2017). It has grown to be a major issue in online communities since it not only stifles positive and healthy discourse but also puts those exposed to it at risk for emotional and psychological harm.

Nowadays, people use different kinds of content on the social media platform, including video, audio, memes and text, to share their opinion or interact with other people (Das and Mukherjee, 2023; Das et al., 2023). The complexity of the computational processing of social media is more for multimodal data, which includes video, audio and text modalities because of the ambiguity at various levels as these types of data are more user-oriented and contextual(Schreck and Keim, 2012).

While there has been a study on abusive language identification and sentiment analysis in the English language for multimodal data, there is a significant lack of exploring these subjects specifically in the context of Dravidian languages. The situation for countries like India is more complicated due to the immense language diversity [1]. Tamil, Telugu, Malayalam and Kannada are Dravidian languages(Krishnamurti, 2003) that are largely spoken in southern India and have a rich linguistic heritage. However, the limited examination of abusive language detection and sentiment analysis in these languages presents unique obstacles and potential for research. As part of this shared task, we have explored Tamil and Malayalam languages(Chakravarthi et al., 2021c,a; Premjith et al., 2022). This shared task on multimedia social media analysis in Dravidian languages includes two sub-tasks -

[1] https://en.wikipedia.org/wiki/Languages_of_India

Third Workshop on Speech and Language Technologies for Dravidian Languages,
pages 217–224, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-085-4_031

1. **Multimodal detection of abusive content in Tamil Language**: This sub-task involves developing models that can analyze textual, speech and visual components of videos from social media platforms, such as YouTube and predict whether they are *abusive* or *non-abusive*(Castro et al., 2019).

2. **Multimodal sentiment analysis in Dravidian languages[Tamil and Malayalam]**: This sub-task involves developing models that can analyze textual, speech and visual components of videos in Tamil and Malayalam from social media platforms, such as YouTube and identify the sentiments expressed in them. The videos are labeled into five categories: *highly positive*, *positive*, *neutral*, *negative* and *highly negative*.

The analysis of multi-modalities has gained significant importance, especially in the realm of video data, which encompasses various modalities such as video frames, speech signals and text transcripts. When training a machine learning model for sentiment analysis, it becomes crucial to incorporate features from these three modalities. Our research specifically targets abusive video detection, utilizing multiple modalities. The primary objective is to identify and remove hateful content from social networks. By considering the combined information from video frames, speech signals and text transcripts, our approach aims to effectively detect and mitigate abusive and harmful content circulating within these platforms.

The paper outlines the methodologies we employed to identify abusive content in Tamil, as well as perform sentiment analysis in Tamil and Malayalam languages on the shared task "Multimodal Abusive Language Detection and Sentiment Analysis in Dravidian Languages: DravidianLangTech@RANLP 2023"(Ashraf et al., 2021; Chakravarthi et al., 2021a; B et al., 2023; Chakravarthi et al., 2021c). To extract text features, we employed the transformer-based model mBERT, while for image feature extraction, we utilized the pre-trained ViT Model. Additionally, MFCC was employed for audio feature extraction. These approaches proved successful, leading us to secure the coveted **first** place in the final leaderboard standings for both tasks. Our techniques and models demonstrate the effectiveness of utilizing these feature extraction methods in the context of Dravidian languages.

2 Related Work

It is vital to filter the abusive content and inflammatory material that is constantly being posted on social media platforms. However, manual screening is nearly difficult due to the overwhelming volume of incoming posts. The research community gave this problem a lot of attention. According to numerous studies, posts in various languages on social media platforms are likely to be insulting or hateful. However, the majority of them spoke only English. There hasn't been much work done to address these concerns in Dravidian languages. This section discusses some of the Multimodal abusive language detection and sentiment analysis methods and briefly explains the multi-modal techniques used so far to detect abusive language.

2.1 Multi-modal abusive language detection

Most of the abusive language detection research were carried out considering textual or Image information (Li, 2021; Mandl et al., 2021; Ghanghor et al., 2021; Suryawanshi et al., 2020; Yasaswini et al., 2021; Chakravarthi et al., 2021a; Andrew, 2021) . There is very less work on Tamil language for detecting abusive content due to lack of resources in this language. As per the shared task we have video ,audio and text data for carrying out the research (Chakravarthi et al., 2021c,a). There are almost no work on multi-modal hate speech detection on Tamil language. Though there many researches on multi-modal hate speech detection for different languages other than Tamil (Das et al., 2023; Thapa et al., 2022; Das et al., 2021a) where they have considered video ,audio and text feature. Such multi-modal schemes typically use unimodal methods like CNNs, LSTMs or BERT to encode text and deep learning models such as ResNet, InceptionV3 to encode images and then perform multi-modal fusion using simple concatenation, gated summation, bi-linear trans-formation, or attention-based methods. Multi-modal bi-transformers like ViLBERT and Visual BERT have also been applied(Kiela et al., 2020).

As part of abusive speech detection, an array of techniques with diverse architecture ranging from video-based, text-based model, image-based model and multi-modal models have been employed(Mozafari et al., 2020; Das et al., 2023).

2.2 Multi-modal sentiment analysis in Dravidian languages

Multimodal sentiment analysis has attracted more and more attention recently (Baltrušaitis et al., 2019; Soleymani et al., 2017; Premjith et al., 2022). Most of the sentiment analysis based on audio or Text in English language (Poria et al., 2018, 2019). But in Dravidian Languages there are almost no work on multimodal sentiment analysis due to lack of resources and study in this area still seems to be in its infancy for this language . We have few woks in Dravidian languages but most of them are based on Text or audio (Ou and Li, 2020; Chakravarthi et al., 2021b). A lot of research has concentrated on creating a novel fusion network based on this topology to better capture multimodal representation(Cambria et al., 2018; Williams et al., 2018; Sahay et al., 2020; Blanchard et al., 2018). As per the shared task of sentiment analysis, we have video, audio and text data for carrying out the research (Chakravarthi et al., 2021c,a). Multimodal sentiment analysis mainly focuses on utilizing multiple resources to predict human emotions. Most multimodal models focus on three modalities: acoustic, visual and text; thus, we also experiment with multimodal sentiment analysis in Tamil and Malayalam languages, where we leverage all three modalities – text, audio and video. The videos are labeled into five categories: *highly positive*, *positive*, *neutral*, *negative* and *highly negative*.

3 Dataset Description

The competition organizers have released data sets for two different languages, Tamil and Malayalam (Chakravarthi et al., 2021c).

However, for the abusive language classification shared task, the dataset was only released for the Tamil language. Competition organizers have provided us with Video, Audio and the extracted texts present in them. The train, dev and test set distributions for both of them are as follows in Table 1.

Split	Abusive	Non-Abusive	Total
Train	38	32	70
Test	9	9	18

Table 1: Offensive Language Dataset Distribution (Tamil)

For the sentiment analysis shared task, the datasets were released for both Tamil and Malayalam languages. Sentiments are labeled into five categories for each language: *highly positive*, *positive*, *neutral*, *negative* and *highly negative*. The train, dev and test set distributions for both Tamil and Malayalam languages are showed as follows in the below Table 2.

Category	Tamil			Malayalam		
	Train	Test	Dev	Train	Test	Dev
highly positive	5	1	2	5	2	2
positive	29	5	4	31	3	5
neutral	4	2	2	5	2	1
negative	3	1	1	8	2	2
highly negative	3	1	1	1	1	0
Total	44	10	10	50	10	10

Table 2: Sentiment analysis Dataset Distribution (Tamil & Malayalam)

4 Methodology

In this section, we discuss the different parts of the pipeline that we pursued for the detection of *abusive* or *non-abusive* Language for the Tamil language using the dataset. Initially, we explored the visual aspects of the videos. Subsequently, the textual information is considered and used transformer-based pre-trained model mBERT. Then we considered audio-based MFCC features for the modeling. Finally, the visual, audio and textual features are combined to make more robust abusive content classification and sentiment analysis. Along with this, we will also discuss sentiment analysis in Tamil and Malayalam languages task.

4.1 Problem Formulation:

Task 1: Abusive Language Detection (Binary Classification) for Tamil : We formulate the abusive video detection problem in this paper as follows. Given a video V, the task can be represented as a binary classification problem. Each video is to be classified as abusive (y = 1) or non-abusive (y = 0). A video V can be expressed as a sequence of frames, i.e., F ={f1, f2, .., fn}, the associated audio A and the extracted video transcript T = {w1, w2, ..., wm}, consisting of a sequence of words. We aim to learn such a hate video classifier Z : Z(F; A; T) $\rightarrow$ y, where y belongs to {0, 1} is the ground-truth label of a video.

Task 2: Sentiment Analysis (Multi-class Classification) for Tamil and Malayalam: Given a video V, the objective is to classify its sentiment into one of five categories, denoted by S = {0, 1, 2,

3, 4}, representing *highly positive*, *positive*, *neutral*, *negative* and *highly negative* sentiments, respectively. The video V can be expressed as a sequence of frames, denoted as F = {f1, f2, ..., fn}. It also contains associated audio, denoted as A and an extracted video transcript T = {w1, w2, ..., wm}, consisting of a sequence of words in the specific language. The sentiment classifier is defined as S(F; A; T) $\rightarrow$ y, where y belongs to S represents the ground-truth label of the video, indicating the sentiment category it belongs to.

We have followed the below-mentioned methods for both **Task 1 (Abusive Language Detection)** and **Task 2 (Sentiment Analysis)** . For Task 1, we have done the modeling for the Tamil language; for Task 2, we have done similar modeling for both Tamil and Malayalam Languages separately. Along with it, as the data was very less for sentiment analysis for both languages, hence we have merged the data set for both languages and performed only for Fusion Model.

4.2 Uni-modal Models

As part of our initial experiments, we created the following three uni-model models, utilizing text features, audio features and image-based features.

mBERT:(Devlin et al., 2019) (multilingual BERT) is a transformer-based language model that has been pre-trained on a large corpus of multilingual text data. It is designed to handle multiple languages and exhibits strong cross-lingual transfer learning capabilities. mBERT captures contextualized representations of words and sentences, enabling it to understand the nuances of different languages and perform well on various natural language processing tasks. With its shared architecture and shared vocabulary, mBERT allows for efficient knowledge transfer between languages, making it a versatile and widely used model for multilingual applications. We pass all the texts associated with the video through the mBERT model and extracted 768-dimensional feature vectors.

Vision Transformer: (Dosovitskiy et al., 2020) The Vision Transformer (ViT) model is a transformer-based architecture specifically designed for computer vision tasks. Unlike traditional convolutional neural networks (CNNs), ViT applies self-attention mechanisms to capture global dependencies in images. It divides the input image into patches and treats them as tokens, allowing the model to learn representations for each patch and their interactions. ViT has shown promising results on various vision tasks, such as image classification, object detection and image generation, demonstrating the power of transformer-based models in the field of computer vision. As our focus is to detect abusive videos or sentiments associated with a video, we cannot use Vision Transformer directly. With the help of OpenCV (Open Source Computer Vision Library), we extracted images from the video for each 1 sec. We uniformly take 30 frames for each video and pass it through the pre-trained Vision Transformer(ViT) (Dosovitskiy et al., 2020) model to get a 768-dimensional feature vector for each frame and finally pass it through the LSTM network to obtain the prediction.

MFCC: The Mel Frequency Cepstral Coefficients (MFCC) (Xu et al., 2004) is one of the widely used techniques for describing audio and research has shown that it is efficient for difficult tasks including lung sound classification(Jung et al., 2021) and speaker identification. We use the MFCC features to obtain a representation of the audio in our dataset. Utilizing the free software Librosa[2], we create a 40-dimensional vector to represent the audio in order to build the MFCC characteristics.

4.3 Fusion Model

The models presented in the preceding subsections are unable to take use of the relationship between the features derived from the various modalities (such as video, audio and text transcript). We try to substantially merge the text-based, audio-based and vision-based models in order to harness the benefits of all the modalities effectively. For Task 1 and Task 2, in particular, we build the following models – **(mBERT + ViT + MFCC)**, **+** refers to the combination operation of the three modalities through a trainable neural network (aka fusion layer). We denote this model as **Fusion 1.**

Due to very less dataset for Task 2, along with the above-mentioned Fusion approach **(Fusion 1)** we have merged data set for both languages and build same Fusion Model - **(mBERT + ViT + MFCC)** as mentioned previously. We denote this model as **Fusion 2**.

All the models are trained with cross-entropy loss functions and Adam optimizer for 30 epochs.

In both Binary-class classification and Multi-class classification, data is imbalanced. To balance the data, an extensive study has been conducted in

[2]https://librosa.org/doc/latest/index.html

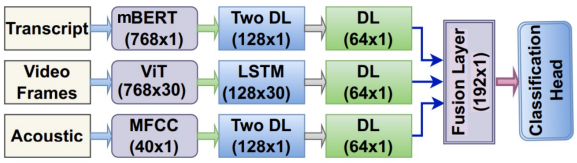

Figure 1: Illustrates the overall modeling pipeline

this area. Both oversampling and under-sampling are widely used techniques for balancing data, although they both have clear drawbacks. Using the class weight procedure, we attempted to reduce the effect of data imbalance.

5 Results

For **Task 1**, we observe among the uni-modal models and fusion-based model, mBERT and MFCC have the highest weighted F1 score of **0.5786** (mBERT: 0.5786, ViT: 0.5555, MFCC: 0.5786, Fusion (mBERT + ViT + MFCC): 0.5555). Table 3 demonstrates the performance of each model.

Abusive Language Detection- Tamil		
Model	**Accuracy**	**F1 Score(w)**
MFCC	**0.611111**	**0.578595**
mBERT	**0.611111**	**0.578595**
ViT	0.555556	0.555556
Fusion	0.555556	0.555556

Table 3: Performance Comparisons of Each Model.w: Weighted-Average. The best performance in each column is marked in **bold** and second best is underlined

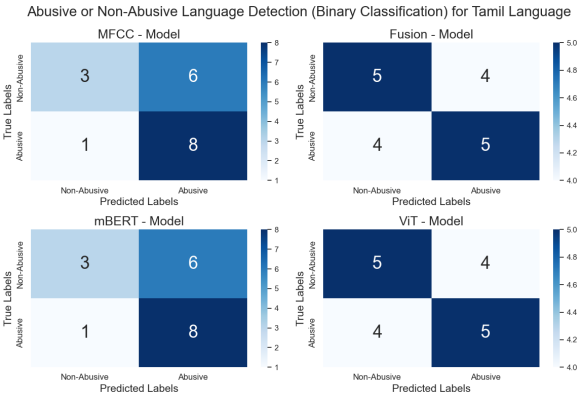

Figure 2: Confusion Matrix on Test Data for Each Model

For **Task 2 -Tamil language**, we observe among the uni-modal models and Fusion Models, ViT has highest weighted F1 score of **0.357** (mBERT: 0.250, ViT: 0.357, MFCC: 0.272, Fusion 1 (BERT + ViT + MFCC): 0.307 , Fusion 2 (BERT + ViT + MFCC): 0.272).

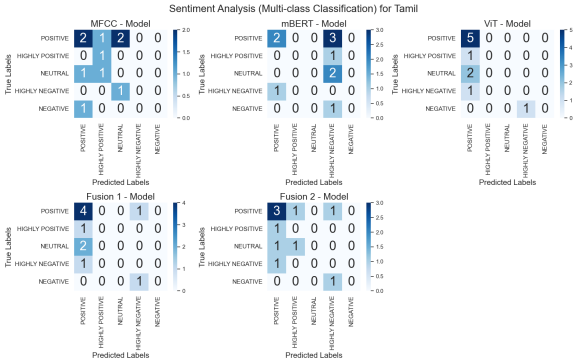

Figure 3: Sentiment Analysis (Multi-class Classification) for Tamil

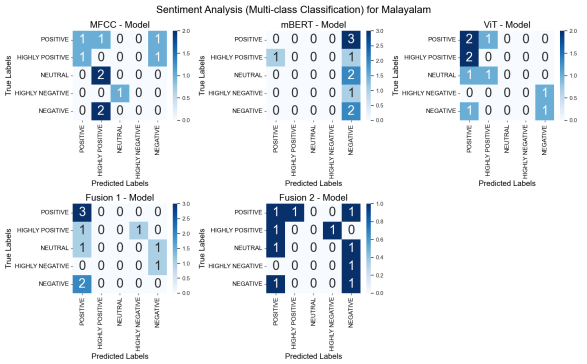

Figure 4: Sentiment Analysis (Multi-class Classification) for Malayalam

For **Task 2 -Malayalam language**, we observe among the uni-modal models and Fusion Models, ViT has highest weighted F1 score of **0.233** (mBERT: 0.0727, ViT: 0.233, MFCC: 0.120, Fusion 1 (BERT + ViT + MFCC): 0.180 , Fusion 2 (BERT + ViT + MFCC): 0.152).

To further understand the model's weakness, we show the confusion matrix of each model in Figure 2, 3 and 4 . We observe that both MFCC & BERT model performs better on the abusive language data, ViT performs better on the sentiment analysis data points for Tamil and Malayalam language. Insufficient data is the main reason behind poor performance on sentiment analysis for both languages and data is highly imbalanced for each class.

6 Conclusion

In this shared task, we deal with a novel problem of detecting Tamil abusive language and Sentiment analysis for both Tamil and Malayalam language. We evaluated different uni-modal models and in-

	Sentiment Analysis - Tamil			Sentiment Analysis - Malayalam		
Model	**Accuracy**	**F1 Score(w)**	**F1 Score(m)**	**Accuracy**	**F1 Score(w)**	**F1 Score (m)**
MFCC	0.3	0.272222	0.188889	0.1	0.120000	0.080000
mBERT	0.2	0.250000	0.100000	0.2	0.072727	0.072727
ViT	**0.5**	**0.357143**	**0.142857**	**0.3**	**0.233333**	**0.188889**
Fusion 1	0.4	0.307692	0.123077	0.3	0.180000	0.120000
Fusion 2	0.3	0.272727	0.109091	0.2	0.152381	0.123810

Table 4: Performance Comparisons of Each Model.w: Weighted-Average. m: macro, The best performance in each column is marked in bold and the second best is underlined

troduced a fusion model. We found that text-based model mBERT and Audio based MFCC performs better on the abusive language classification. For the sentiment analysis task, the video-based unimodal model ViT performs better on the sentiment analysis data points for Tamil and Malayalam languages. We plan to explore further other vision-based models to improve performance as an immediate next step.

References

Judith Jeyafreeda Andrew. 2021. JudithJeyafreedaAndrew@DravidianLangTech-EACL2021:offensive language detection for Dravidian code-mixed YouTube comments. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 169–174, Kyiv. Association for Computational Linguistics.

Noman Ashraf, Arkaitz Zubiaga, and Alexander Gelbukh. 2021. Abusive language detection in youtube comments leveraging replies as conversational context. *PeerJ Computer Science*, 7:e742.

Premjith B, Sowmya V, Jyothish Lal G, Bharathi Raja Chakravarthi, Nandhini K, Rajeswari Natarajan, Abirami Murugappan, Bharathi B, Kaushik M, Prasanth S.N, Aswin Raj R, and Vijai Simmon S. 2023. Findings of the Multimodal Abusive Language Detection and Sentiment Analysis in Dravidian Languages @ dravidianlangtech-ranlp 2023. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages DravidianLangTech 2023*. Recent Advances in Natural Language Processing.

Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2019. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443.

Somnath Banerjee, Maulindu Sarkar, Nancy Agrawal, Punyajoy Saha, and Mithun Das. 2021. Exploring transformer based models to identify hate speech and offensive content in english and indo-aryan languages. *arXiv preprint arXiv:2111.13974*.

Nathaniel Blanchard, Daniel Moreira, Aparna Bharati, and Walter J Scheirer. 2018. Getting the subtext without the text: Scalable multimodal sentiment classification from visual and acoustic modalities. *arXiv preprint arXiv:1807.01122*.

Erik Cambria, Devamanyu Hazarika, Soujanya Poria, Amir Hussain, and RBV Subramanyam. 2018. Benchmarking multimodal sentiment analysis. In *Computational Linguistics and Intelligent Text Processing: 18th International Conference, CICLing 2017, Budapest, Hungary, April 17–23, 2017, Revised Selected Papers, Part II 18*, pages 166–179. Springer.

Santiago Castro, Devamanyu Hazarika, Verónica Pérez-Rosas, Roger Zimmermann, Rada Mihalcea, and Soujanya Poria. 2019. Towards multimodal sarcasm detection (an _obviously_ perfect paper). *arXiv preprint arXiv:1906.01815*.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Navya Jose, Thomas Mandl, Prasanna Kumar Kumaresan, Rahul Ponnusamy, RL Hariharan, John Philip McCrae, Elizabeth Sherly, et al. 2021a. Findings of the shared task on offensive language identification in tamil, malayalam, and kannada. In *Proceedings of the first workshop on speech and language technologies for Dravidian languages*, pages 133–145.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, Elizabeth Sherly, John P McCrae, Adeep Hande, Rahul Ponnusamy, Shubhanker Banerjee, et al. 2021b. Findings of the sentiment analysis of dravidian languages in code-mixed text. *arXiv preprint arXiv:2111.09811*.

Bharathi Raja Chakravarthi, KP Soman, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Kingston Pal Thamburaj, John P McCrae, et al. 2021c. Dravidianmultimodality: A dataset for multi-modal sentiment analysis in tamil and malayalam. *arXiv preprint arXiv:2106.04853*.

Mithun Das, Somnath Banerjee, and Punyajoy Saha. 2021a. Abusive and threatening language detection in urdu using boosting based and bert based models: A comparative approach. *arXiv preprint arXiv:2111.14830*.

Mithun Das, Binny Mathew, Punyajoy Saha, Pawan Goyal, and Animesh Mukherjee. 2020. Hate speech in online social media. *ACM SIGWEB Newsletter*, (Autumn):1–8.

Mithun Das and Animesh Mukherjee. 2023. Transfer learning for multilingual abusive meme detection. In *Proceedings of the 15th ACM Web Science Conference 2023*, pages 245–250.

Mithun Das, Rohit Raj, Punyajoy Saha, Binny Mathew, Manish Gupta, and Animesh Mukherjee. 2023. Hatemm: A multi-modal dataset for hate video classification. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 1014–1023.

Mithun Das, Punyajoy Saha, Ritam Dutt, Pawan Goyal, Animesh Mukherjee, and Binny Mathew. 2021b. You too brutus! trapping hateful users in social media: Challenges, solutions & insights. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*, pages 79–89.

J. Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL*.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Nikhil Ghanghor, Parameswari Krishnamurthy, Sajeetha Thavareesan, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2021. IIITK@DravidianLangTech-EACL2021: Offensive language identification and meme classification in Tamil, Malayalam and Kannada. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 222–229, Kyiv. Association for Computational Linguistics.

SY Jung, CH Liao, YS Wu, SM Yuan, and CT Sun. 2021. Efficiently classifying lung sounds through depthwise separable cnn models with fused stft and mfcc features. diagnostics 2021, 11, 732.

Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Pratik Ringshia, and Davide Testuggine. 2020. The hateful memes challenge: Detecting hate speech in multimodal memes. *Advances in Neural Information Processing Systems*, 33:2611–2624.

Bhadriraju Krishnamurti. 2003. *The dravidian languages*. Cambridge University Press.

Zichao Li. 2021. Codewithzichao@DravidianLangTech-EACL2021: Exploring multimodal transformers for meme classification in Tamil language. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 352–356, Kyiv. Association for Computational Linguistics.

Thomas Mandl, Sandip Modha, Anand Kumar M, and Bharathi Raja Chakravarthi. 2021. Overview of the hasoc track at fire 2020: Hate speech and offensive language identification in tamil, malayalam, hindi, english and german. In *Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation*, FIRE '20, page 29–32, New York, NY, USA. Association for Computing Machinery.

Marzieh Mozafari, Reza Farahbakhsh, and Noel Crespi. 2020. A bert-based transfer learning approach for hate speech detection in online social media. In *Complex Networks and Their Applications VIII: Volume 1 Proceedings of the Eighth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2019 8*, pages 928–940. Springer.

Xiaozhi Ou and Hongling Li. 2020. Ynu@ dravidiancodemix-fire2020: Xlm-roberta for multi-language sentiment analysis. In *FIRE (Working Notes)*, pages 560–565.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, Florence, Italy. Association for Computational Linguistics.

Soujanya Poria, Navonil Majumder, Devamanyu Hazarika, Erik Cambria, Alexander Gelbukh, and Amir Hussain. 2018. Multimodal sentiment analysis: Addressing key issues and setting up the baselines. *IEEE Intelligent Systems*, 33(6):17–25.

B Premjith, Bharathi Raja Chakravarthi, Malliga Subramanian, B Bharathi, Soman Kp, V Dhanalakshmi, K Sreelakshmi, Arunaggiri Pandian, and Prasanna Kumaresan. 2022. Findings of the shared task on multimodal sentiment analysis and troll meme classification in dravidian languages. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pages 254–260.

Saurav Sahay, Eda Okur, Shachi H Kumar, and Lama Nachman. 2020. Low rank fusion based transformers for multimodal sequences. *arXiv preprint arXiv:2007.02038*.

Tobias Schreck and Daniel Keim. 2012. Visual analysis of social media data. *Computer*, 46(5):68–75.

Mohammad Soleymani, David Garcia, Brendan Jou, Björn Schuller, Shih-Fu Chang, and Maja Pantic. 2017. A survey of multimodal sentiment analysis. *Image and Vision Computing*, 65:3–14.

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).

Surendrabikram Thapa, Aditya Shah, Farhan Jafri, Usman Naseem, and Imran Razzak. 2022. A multimodal dataset for hate speech detection on social media: Case-study of russia-Ukraine conflict. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, pages 1–6, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Zeerak Waseem, Thomas Davidson, Dana Warmsley, and Ingmar Weber. 2017. Understanding abuse: A typology of abusive language detection subtasks. *arXiv preprint arXiv:1705.09899*.

Jennifer Williams, Ramona Comanescu, Oana Radu, and Leimin Tian. 2018. Dnn multimodal fusion techniques for predicting video sentiment. In *Proceedings of grand challenge and workshop on human multimodal language (Challenge-HML)*, pages 64–72.

Min Xu, Ling-Yu Duan, Jianfei Cai, Liang-Tien Chia, Changsheng Xu, and Qi Tian. 2004. Hmm-based audio keyword generation. In *Pacific-Rim Conference on Multimedia*, pages 566–574. Springer.

Konthala Yasaswini, Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITT@DravidianLangTech-EACL2021: Transfer learning for offensive language detection in Dravidian languages. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 187–194, Kyiv. Association for Computational Linguistics.