Multimodal Hate Speech Detection in Dravidian languages: DravidianLangTech@NAACL 2025

* Indicates required question

Email address (Please use the same email id for each submission) *

vasantharank.ncc@gmail.com

What is your team's name (this will appear in the paper. Please use the same name for each submission) *

The Deathly Hallows

What is your CodaLab username? *

VASANTHARAN

Please list the names and emails of all the team members (separated by comma) *
E.g: John Davis (johnd@gmail.com), Emma Watson (emmaw@gmail.com)

VASANTHARAN K (vasantharank.ncc@gmail.com), PRETHISH GA (prethish0409@gmail.com)

Briefly describe the method you used in your submission (this might appear in task description paper) *

In this submission, we implemented a multimodal approach leveraging both audio and text features for classification tasks. For audio data, we applied augmentation techniques including noise addition, time-stretching, and pitch-shifting to enhance data diversity. Features were extracted using MFCCs after normalization to ensure consistent scaling and dynamics. A convolutional neural network (CNN) was employed for audio classification, comprising stacked Conv2D, BatchNormalization, MaxPooling2D, and Dropout layers to improve generalization and performance. For text data, we utilized the "xlm-roberta-large" pre-trained transformer model to generate embeddings. Text preprocessing included tokenization and removal of stopwords to optimize inputs. Extracted embeddings were fed into a feed-forward neural network (FNN) with dense layers, BatchNormalization, and Dropout for classification. The combination of augmentation, pre-trained models, and robust architectures allowed effective handling of multimodal data, ensuring improved accuracy and generalization. Both pipelines were trained with categorical cross-entropy loss and Adam optimizer for efficient convergence.

Did you use any external dataset? *

- ○ Yes
- ◉ No

If your answer to the previous question is Yes, please give the name and link for the dataset, otherwise write NA. *

NA

What is the most interesting aspect of your system? (If we are to briefly describe your system in the task description paper, what would you want to mention?) *

The most interesting aspect of our system is its seamless integration of multimodal data—audio and text—enhanced through sophisticated augmentation and pre-trained transformer embeddings. For audio, diverse augmentation techniques like noise addition, time-stretching, and pitch-shifting enrich the dataset, while MFCC extraction ensures robust feature representation. For text, leveraging "xlm-roberta-large" embeddings enables deep contextual understanding across languages. The use of specialized neural architectures—CNN for audio and FNN for text—optimized with BatchNormalization and Dropout ensures efficient learning and generalization. This unified approach demonstrates adaptability and effectiveness in handling diverse modalities, achieving high accuracy and robustness in classification tasks.

Upload a single zip folder containing the tsv files named with your team's name, task name and language. *

Upload 1 supported file. Max 10 MB.

The Deathly Hall...

If we run this task again next year, what changes/additions would you be interested in seeing?

Try to improve the size of the dataset.

Please provide any further changes or comments

Your answer

Submit Clear form

Never submit passwords through Google Forms.