Overview of Memotion 3: Sentiment and Emotion Analysis of Codemixed Hinglish Memes

Shreyash Mishra*[1], S Suryavardan*[1], Megha Chakraborty[2], Parth Patwa[3], Anku Rani[2], Aman Chadha†[4,5], Aishwarya Reganti[6], Amitava Das[2], Amit Sheth[2], Manoj Chinnakotla[7], Asif Ekbal[8] and Srijan Kumar[9]

[1]*IIIT Sri City, India*
[2]*University of South Carolina, USA*
[3]*UCLA, USA*
[4]*Stanford University, USA*
[5]*Amazon AI, USA*
[6]*CMU, USA*
[7]*Microsoft, USA*
[8]*IIT Patna, India*
[9]*Georgia Tech, USA*

Abstract

Analyzing memes on the internet has emerged as a crucial endeavor due to the impact this multi-modal form of content wields in shaping online discourse. Memes have become a powerful tool for expressing emotions and sentiments, possibly even spreading hate and misinformation, through humor and sarcasm. In this paper, we present the overview of the Memotion 3 shared task, as part of the DeFactify 2 workshop at AAAI-23. The task released an annotated dataset of Hindi-English code-mixed memes based on their Sentiment (Task A), Emotion (Task B), and Emotion intensity (Task C). Each of these is defined as an individual task and the participants are ranked separately for each task. Over 50 teams registered for the shared task and 5 made final submissions to the test set of the Memotion 3 dataset. CLIP, BERT modifications, ViT etc. were the most popular models among the participants along with approaches such as Student-Teacher model, Fusion, and Ensembling. The best final F1 score for Task A is 34.41, Task B is 79.77 and Task C is 59.82.

Keywords

Memes, codemixed, multimodal, Hindi-English

1. Introduction

The term *meme* is derived from the Ancient Greek word *mimema*, meaning *imitated thing*, which comes from the verb *mimeisthai*, meaning *to mimic*. The term was coined by Richard Dawkins, a British evolutionary biologist, in his book *The Selfish Gene* [1]. Dawkins used the concept of a

*Equal contribution.
†Work does not relate to position at Amazon.

De-Factify 2: 2nd Workshop on Multimodal Fact Checking and Hate Speech Detection, co-located with AAAI 2023
shreyash.m19@iiits.in (S. Mishra*); suryavardan.s19@iiits.in (S. Suryavardan*); amitava@mailbox.sc.edu (A. Das)

meme to explain the spread of cultural phenomena and ideas, drawing parallels between memes and genes. Dawkins provided examples of memes such as melodies, catchphrases, fashion, and arch-building technology in his book. Interestingly, the word *meme* is a self-describing term, also known as autological, meaning that it is a meme itself.

The reach of online content is vast and has the capacity to reach a massive viewership. Memes propagate through multiple mediums including social media, messaging, apps, and emails. An article on memes quotes "When you plant a fertile meme in my mind, you literally, parasitize my brain" [2]. This quote indicates the impact of memes on a viewer's mind. It spreads to others who find it motivating, offensive, amusing, or relatable in some ways [3]. Memes are used to convey an opinion. When a person comes across a meme that they find relatable, they often share it with others who they think would find relatable. Different people interpret memes differently which means what can be offensive for one person might not be for other. This difference in interpretation is a result of cultural, gender, and demographic differences.

The memes can also be used to spread offensive and hateful content [4]. Identifying and halting the dissemination of hateful memes is an arduous undertaking for both human beings and AI models, as it requires comprehending the intricate subtleties and socio-political contexts that form their interpretation. The deficiencies of current hate speech moderation techniques highlight the pressing need to enhance the effectiveness of automated hate speech detection. Given their subtlety and multi-modal nature, memes pose a more challenging issue to address compared to only text.

In this paper, we present the findings of the Memotion 3 shared task, where participants were provided with an annotated dataset of 10k memes [5], and were tasked with detecting the sentiment, emotion and emotion intensity of the meme. Unlike the previous iteration of memotion [6, 7] which provided English memes, current iteration studies codemixed Hindi-English (Hinglish) memes.

The paper is organized as follows: we describe the related work and the task in section 2 and 3 respectively; Section 4 contains the details of the dataset we collected for Memotion analysis: Memotion 3.0; followed by a brief description of baseline models and their results in 5. We conclude and mention some possible future works in section 6.

2. Related Work

Sentiment and emotion analysis: There has been significant research on sentiment analysis for text over many years [8]. Work on sentiment analysis using ML methods like SVM, logistic regression, random forest, XGBoost, k-nearest neighbor has been done in [9, 10, 11]. Works which use DL methods include [12, 13, 14]. For detailed surveys on sentiment analysis in social media, please refer to [15, 16].

The HaHa shared task provides a dataset for humor detection on social media [17]. Öhman et al. [18] release an English dataset to detect eight emotions like joy, sadness, disgust etc. A comprehensive survey of textual emotion detection is provided in [19]

Hatespeech detection: It is important to detect hatespeech to keep social media safe for everyone including minorities. Towards this goal, researchers have curated and released annotated datasets [20]. The offenseval shared task [21] at SemEval 2019 releases an annotated

dataset of 14 English tweets to detect the type and target of offensive language. The HatEval shared task releases English and Spanish datasets to detect hate towards women and immigrants. Patwa et al. [22] conduct a shared task on Hindi hostile tweets detection. The TRAC workshop series [23, 24, 25] conducts multiple shared tasks to detect aggression and misogyny in English, Hindi, and Bengali datasets.

Methods to detect hatespeech in text include CNNs and RNNS [26, 27, 28, 29], Bert-like models [30, 31, 32], incorporating linguistic characteristics [33] etc.

Codemixed Language Processing : Codemixed language processing is a challenging task because the informal mixing of 2 or more languages and the proliferation of unique number of ways to write the same word [34, 35]. The Sentimix task [36] at semeval 2020 focused on sentiment analysis of Hinglish and Spanish-English tweets. [37] organized a shared task on detecting offense in 3 codemixed dravidian languages - Tamil [38], Malayalam[39] and Kannada[40]. Methods explored to tackle codemixing include graph convolutional netowrks [41], BERT based models [42, 43], modifying loss function [44, 45] or positional embeddings [46] to incorporate codemixing etc.

Multimodal analysis : Although most of the existing research focuses on unimodal (text) analysis, the use of multi-modal content likes memes and videos is fast increasing. Multimodal datatsets having text and image, or videos are useful for tasks like image captioning, hatespeech detection, emotion analysis in videos, sentiment analysis [47, 48, 49, 50] etc among other tasks. The Hateful Memes Challenge dataset [51] and the multioff dataset [4] address hatespeech and offense detection in memes. However, they are binary classification task and the memes are in English whereas memotion 3 has multi-class and multi-label tasks on code-mixed data. There have been very few works on code-mixed meme analysis. [52] conduct a shared task to detect trolling in Tamil codemixed memes whereas [53] release a dataset to detect hate in codemixed Bengali memes. Popular methods for multimodal learning include image-text joint embeddings [54, 55, 56] and transformer based models [57, 58].

Previous iteration of Memotion : Memotion 1 [6] and Memotion 2[7] shared task released datasets of 10k memes each [6, 59]. These datasets were annotated on the same tasks as Memotion 3. However, both these datasets only focus on English memes, where as in memotion 3, we focus on Hinglish codemixed memes. Methods like ensembling [60, 61, 62] and bert-like models [63, 64, 65] were common across memotion 1 and Memotion 2.

3. Task Details

3.1. Tasks

Memotion 3 is the 3rd iteration of the Memotion task. The challenge consists of three sub-tasks:

1. **Task A - Sentiment analysis of memes**: Given a meme, the system is supposed to classify the meme's sentiment as positive, negative or neutral.
2. **Task B - Overall emotion analysis of memes**: This task's goal is to pinpoint certain emotions connected to a given meme. Whether a meme is humorous, sarcastic, offensive, or motivating should be indicated by the system. There are multiple categories in which a meme can fit.

3. **Task C - Classifying the intensity of meme emotions**: The task is to determine the degree to which a particular emotion is being expressed. The ranking of these emotions is as follows:
    a) Humour: Not funny, funny, very funny and hilarious
    b) Sarcasm: Not Sarcastic, little sarcastic, very sarcastic and extremely sarcastic
    c) Offensive: Not offensive, slightly offensive, very offensive and hateful offensive
    d) Motivation: Not motivational, motivational

3.2. Dataset

The tasks were conducted on the Memotion 3 dataset [5]. It consists of Hindi-English codemixed memes, which were collected from selenium based web crawler and they were gathered from various public platforms like Reddit, Google Images, etc and annotated manually. The dataset consists of 10,000 meme images divided into a train-val-test split of 7000-1500-1500. Each meme is annotated for its sentiment, emotion and intensity of emotion. Images also have their corresponding OCR text extracted with the help of Google Vision APIs and their respective URLs. For more details of the dataset, please refer to [5].

3.3. Evaluation

As mentioned previously, there are three tasks. Scoring is done for each task separately, and separate leaderboards are generated. For each task, we use weighted average F1 score to measure the performance of a model. The participants had access to only train and validation set. They were asked to submit a maximum of 3 submissions on the test set for each task, the best of which was selected as part of the leaderboard.

3.4. Baselines

For multi-modal data, it is crucial to take into account both the visual and textual properties, particularly in the case of memes where the context can only be recorded using a mix of both elements. For textual features, we employ a multilingual form of BERT, namely Hinglish-BERT (`BERT-base-multilingual-cased`) [66], which is tuned on Hinglish data. The trained Vision Transformer (ViT) model provides the visual properties [67]. The Hinglish-BERT embedding is concatenated with the pooled output from the ViT model. After passing through an MLP, the combined features are then categorised in a final classification layer. For more details about the baseline, please refer [5].

4. Participating systems

There were 47 team registrations for the task in the Memotion 3.0 task page, of which 5 teams made submissions for the final test set of the dataset. The results for all three tasks are given in the following section and an overview of the 4 teams that presented their description papers are provided below.

Rank	Team	F1 - Scores
1	**NUAA-QMUL-AIIT**	**34.41%**
2	NYCU_TWO	34.20%
3	CUFE	33.77%
4	CSECU-DSG	33.34%
5	Baseline	33.28%
6	wentaorub	32.88%

Table 1
Leaderboard of teams on Task A: Sentiment Analysis.

wentaorub [68] use CLIP [69] for individual text and image embeddings, before concatenating them and passing them through multi-headed attention layers for classification. They also use the OSCAR [70] model in this approach and an ensemble of their models are used for the final submission. Datasets such as Facebook Hateful memes [51], MMHS150k [49] etc. are used for pre-training. This architecture helped them achieve the best performance in Task B and C.

NYCU_TWO [71] propose a two model pipelines, namely Coopoerative Teaching Model (CTM) for task A and Cascaded Emotion Classifier (CEC) for task B and C. A fusion of multi-modal embeddings from pre-trained Swin-Transformer [72] and CLIP are passed to the CTM and CEC pipelines. CEC helps leverage task C predictions for task B by jointly training the model. This team attained the best results in 3 out of the 4 labels in Task B.

NUAA-QMUL-AIIT [73] refer to their approach as Squeeze-and-Excitation Fusion or SEFusion. The textual features from pre-trained RoBERTa [74] and visual features from CLIP-ViT [75] are fused to obtain multi-modal embeddings. The fusion is the SEFusion module, which uses a learned activation of the squeezed features, allowing for weighted fusion of multi-modal embeddings. This approach led to NUAA-QMUL-AIIT being the 1st ranked team in Task A.

CUFE used a LightGBM [76] classifier for classification on every individual emotion or label in all Tasks. The inputs to the classifier were pre-trained Hinglish-DistilBERT for text embeddings and ResNet18 [77] for image embeddings. Other features, such as occurrence of characters, word count etc. from text and number of faces in the memes (extracted using Facenet's [78] multitask cascaded CNN), were also used. CUFE obtained the highest score in 2 labels in Task B and 1 label in Task C.

5. Results

The performance in Task A i.e. sentiment classification is presented in Table 1. Our proposed baseline achieved a score of 33.28%. Out of the five final submissions, four teams managed to surpass the baseline. The top two teams, namely, NUAA-QMUL-AIIT [73] and NYCU_TWO [71] improved on the baseline by 1.1% and 0.9% respectively. The difference in F1 scores for this task is minimal across all participants.

Table 2 shows the leaderboard for Task B. Three teams out-performed the baseline for this task and top performing team `wentaorub` [68] improves on the final score by 5.4%. Team CUFE performs lower than the baseline score by 3.4%, however, they present the highest score on the

Rank	Team	F1 - Scores				
		H	S	O	M	Overall
1	wentorub	88.91	86.74	48.99	**94.44**	**79.77**
2	NYCU_TWO	**89.04**	**86.91**	43.17	**94.44**	78.39
3	NUAA-QMUL-AIIT	87.06	77.97	50.72	**94.44**	77.55
4	BASELINE	84.55	74.82	48.84	90.78	74.74
5	CUFE	82.26	**86.91**	**50.78**	77.60	74.39
6	CSECU-DSG	88.64	86.30	49.32	64.79	72.26

Table 2
Leaderboard of teams on Task B: Emotion analysis {H:Humor, S:Sarcasm, O:Offense, M:Motivation}. All the scores are in percentage. The teams are ranked by their average F1 scores (overall) across all the four emotions. Motivation emotion is the easiest to detect while offense is the most difficult to detect.

Offensive sub-task and on the Sarcasm sub-task, jointly with NYCU_TWO [71]. wentaorub [68], NYCU_TWO, and NUAA-QMUL-AIIT achieved the same score on Motivation, which is 4% higher than the baseline. NYCU_TWO also scored the highest in the Humour sub-task, improving on the baseline by 4.5%. Based on these F1 scores, we can deduce that the Motivation class is the easiest to detect, similar to Memotion 2.0 [59], despite some teams performing poorly on this class. However, in this iteration, the Offensive class is the hardest to detect, instead of the Sarcasm class. No single team performs the best on all the classes.

Rank	Team	F1 - Scores				
		H	S	O	M	Overall
1	wentaorub	**48.37**	50.65	**45.82**	**94.43**	**59.82**
2	NUAA-QMUL-AIIT	46.34	44.29	43.17	**94.43**	57.06
3	CSECU-DSG	42.54	34.41	45.68	93.26	53.97
4	NYCU_TWO	43.28	45.80	44.67	80.57	53.58
5	CUFE	41.66	**51.07**	42.02	77.60	53.09
6	BASELINE	43.03	32.89	42.40	90.78	52.27

Table 3
Leaderboard of teams on Task C: Emotion intensity detection {H:Humor, S:Sarcasm, O:Offense, M:Motivation}. All scores are in percentage. The teams are ranked by average F1 scores (overall) across all the four emotions. Intensity of sarcasm is by far the most difficult to detect. Motivation has only two intensities as opposed to four intensity levels for other emotions.

As shown in the leaderboard for Task C in Table 3, all the participating teams outperform the baseline in this task. The minimum improvement in F1 score is 0.8% by team CUFE and the maximum improvement in the final score is 7.5% by wentaorub [68]. The submission by wentaorub exceeds all other teams in the sub-tasks Humour, Offensive and Motivation. NUAA-QMUL-AIIT [73] matched the highest score of wentaorub in the Motivation category, while CUFE achieved the highest score in the Sarcasm sub-task. The Motivation class has significantly higher F1 scores than other classes, due to it having only 2 intensities whereas other classes have 4 intensities each. The best overall score is only 59.82%, which shows the difficulty of the task.

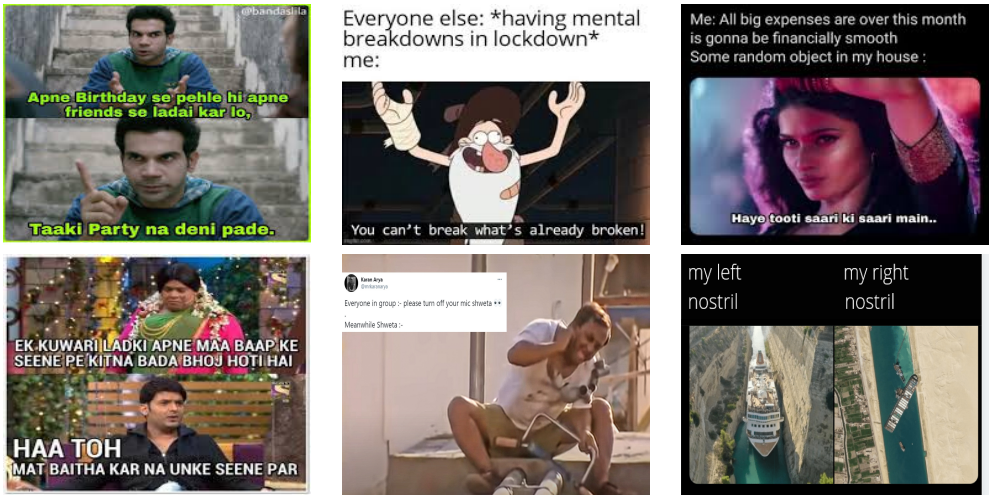

Figure 1: Some samples for Task A from the 121 memes that were predicted incorrectly by all participating teams. Over half of the mis-classified memes are true negatives.

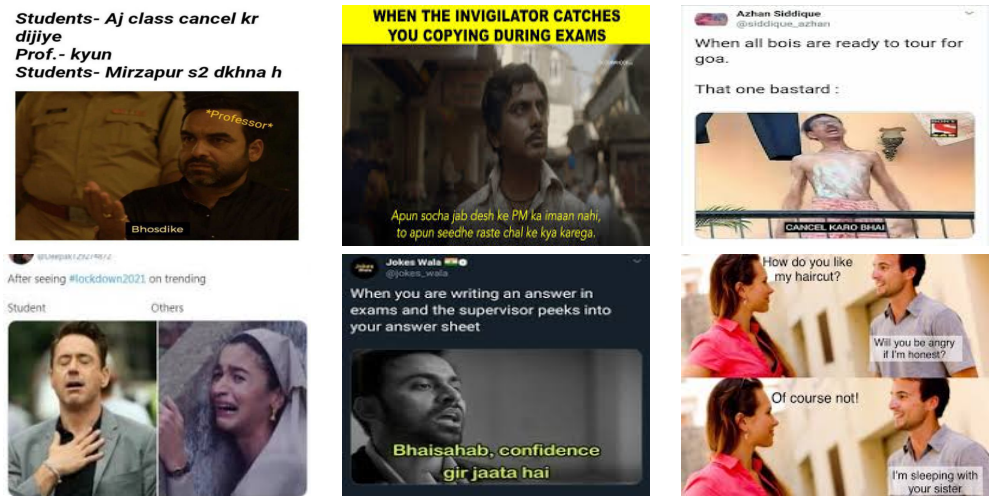

Figure 2: 242 memes from the test set were predicted incorrectly by all participants. Most of such memes are code-mixed and related to humor or sarcasm.

A major observation when comparing performances with the previous iteration of the task [7] is that the performance on Task A is much worse in Memotion 3. Overall Scores in Tasks B and C have remained about the same compared to Memotion 2.

For each individual task, we analyze the memes from the test set that all participants made wrong predictions on. For Task A, there are 121 memes where all participants mis-classified the label, out of which 66 memes were true negative sentiment followed closely by true positive memes. For task B, there are 231 such memes, majority of which belong to "humor" and "sarcasm" class. Finally, for Task C, 421 memes are mis-classified by all the systems - most memes mis-classified by all teams are "Very Funny", "Very Sarcastic", "Slightly Offensive". Some

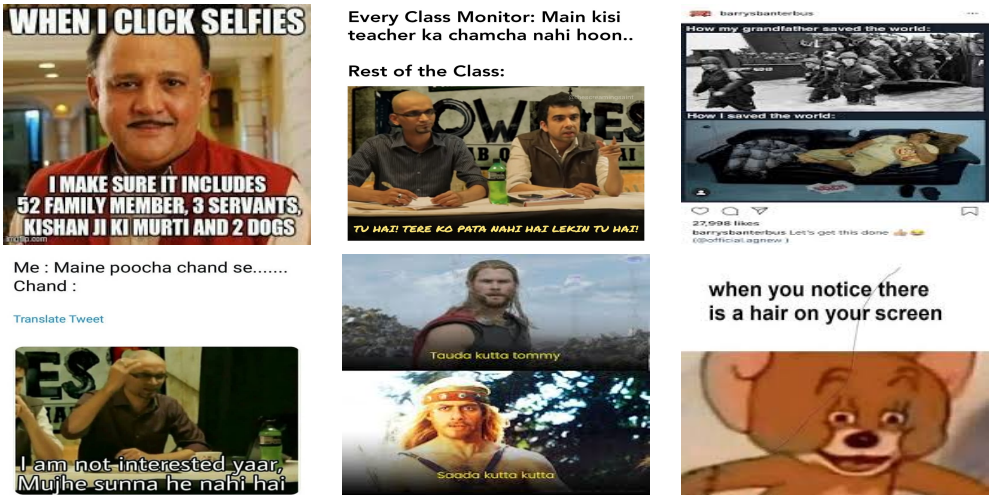

Figure 3: For Task C, none of the teams made correct predictions on over 400 memes from the test set. A large share of these memes is code-mixed, thus indicating the added challenge to the task.

such examples for Task A, B and C are shown in figures 1, 2, and 3 respectively. Further, we note that most of such difficult examples have code-mixed text.

As for the overall performances, only two teams - NUAA-QMUL-AIIT and NYCU_TWO [73, 71] - perform better than the baseline in all tasks.

6. Conclusion and Future Work

In this paper, we summarize the approaches used by the participants for the Memotion 3 task and analyze the results. Due to the multi-modal nature of the dataset, all teams use a pre-trained image and text embedding models. However, each team presents a novel model pipeline. The highest scores achieved in Task A, Task B and Task C of Memotion 3.0 are 34.41%, 79.77% and 59.82% respectively, which shows there is significant room for improvement. On analysis of the results and the mis-classified examples on the test set, we find that "Sarcasm" and "Humour" are difficult to identify, especially in code-mixed memes.

While we address Hind-English code-mixed memes in this paper, future work could include exploring other languages/language pairs. A unified baseline model to analyze memes in multiple languages could also be an interesting possibility.

References

[1] R. Dawkins, The selfish gene, Granada Publishing Lim, 1979.
[2] A. Marwick, Memes, Contexts 12 (2013) 12–13.
[3] N. Akhther, Internet memes as form of cultural discourse: A rhetorical analysis on facebook, 2018. doi:10.31234/osf.io/sx6t7.

[4] S. Suryawanshi, B. R. Chakravarthi, M. Arcan, P. Buitelaar, Multimodal meme dataset (MultiOFF) for identifying offensive content in image and text, in: TRAC, 2020.
[5] S. Mishra, S. Suryavardan, P. Patwa, M. Chakraborty, A. Rani, A. Reganti, A. Chadha, A. Das, A. Sheth, M. Chinnakotla, et al., Memotion 3: Dataset on sentiment and emotion analysis of codemixed hindi-english memes, arXiv preprint arXiv:2303.09892 (2023).
[6] C. Sharma, D. Bhageria, W. Scott, S. PYKL, A. Das, T. Chakraborty, et al., SemEval-2020 task 8: Memotion analysis- the visuo-lingual metaphor!, in: SemEval, 2020.
[7] P. Patwa, S. Ramamoorthy, N. Gunti, S. Mishra, S. Suryavardan, A. Reganti, A. Das, T. Chakraborty, A. Sheth, A. Ekbal, et al., Findings of memotion 2: Sentiment and emotion analysis of memes, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, ceur, 2022.
[8] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in: EMNLP, 2013.
[9] A. I. Saad, Opinion mining on us airline twitter data using machine learning techniques, in: 2020 16th international computer engineering conference (ICENCO), IEEE, 2020.
[10] M. Alzyout, E. A. Bashabsheh, H. Najadat, A. Alaiad, Sentiment analysis of arabic tweets about violence against women using machine learning, in: 12th ICICS, 2021.
[11] E. Prabhakar, M. Santhosh, A. H. Krishnan, T. Kumar, R. Sudhakar, Sentiment analysis of us airline twitter data using new adaboost approach, (IJERT) 7 (2019).
[12] S. T. Kokab, S. Asghar, S. Naz, Transformer-based deep learning models for the sentiment analysis of social media data, Array 14 (2022) 100157.
[13] S. G. Tesfagergish, J. Kapočiūtė-Dzikienė, R. Damaševičius, Zero-shot emotion detection for semi-supervised sentiment analysis using sentence transformers and ensemble learning, Applied Sciences (2022).
[14] K. L. Tan, C. P. Lee, K. M. Lim, K. S. M. Anbananthen, Sentiment analysis with ensemble hybrid deep learning model, IEEE Access 10 (2022) 103694–103704.
[15] L. Yue, W. Chen, X. Li, W. Zuo, M. Yin, A survey of sentiment analysis in social media, Knowledge and Information Systems 60 (2019).
[16] K. Chakraborty, S. Bhattacharyya, R. Bag, A survey of sentiment analysis from social media data, IEEE Transactions on Computational Social Systems 7 (2020) 450–464.
[17] L. Chiruzzo, S. Castro, M. Etcheverry, D. Garat, J. J. Prada, A. Rosá, Overview of haha at iberlef 2019: Humor analysis based on human annotation., in: IberLEF@ SEPLN, 2019.
[18] E. Öhman, M. Pàmies, K. Kajava, J. Tiedemann, Xed: A multilingual dataset for sentiment analysis and emotion detection, 2020. arXiv:2011.01612.
[19] F. A. Acheampong, C. Wenyu, H. Nunoo-Mensah, Text-based emotion detection: Advances, challenges, and opportunities, Engineering Reports (2020).
[20] Z. Waseem, D. Hovy, Hateful symbols or hateful people? predictive features for hate speech detection on Twitter, in: NAACL, 2016.
[21] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, et al., Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval), arXiv:1903.08983 (2019).
[22] P. Patwa, M. Bhardwaj, V. Guptha, G. Kumari, S. Sharma, S. PYKL, A. Das, A. Ekbal, M. S. Akhtar, T. Chakraborty, Overview of constraint 2021 shared tasks: Detecting english covid-19 fake news and hindi hostile posts, in: Combating Online Hostile Posts in Regional Languages during Emergency Situation, Springer International Publishing, 2021, pp. 42–53.

[23] R. Kumar, A. K. Ojha, S. Malmasi, M. Zampieri, Benchmarking aggression identification in social media, in: TRAC workshop, 2018.
[24] R. Kumar, A. K. Ojha, S. Malmasi, M. Zampieri, Evaluating aggression identification in social media, in: TRAC workshop, 2020.
[25] R. Kumar, S. Ratan, S. Singh, E. Nandi, L. N. Devi, et al., The ComMA dataset v0.2: Annotating aggression and bias in multilingual social media discourse, in: LREC, 2022.
[26] B. Gambäck, U. K. Sikdar, Using convolutional neural networks to classify hate-speech, in: Proceedings of the first workshop on abusive language online, 2017, pp. 85–90.
[27] A. Ribeiro, N. Silva, Inf-hateval at semeval-2019 task 5: Convolutional neural networks for hate speech detection against women and immigrants on twitter, in: SemEval, 2019.
[28] K. Winter, R. Kern, Know-center at semeval-2019 task 5: multilingual hate speech detection on twitter using cnns, in: Semeval, 2019.
[29] P. Patwa, S. Pykl, A. Das, P. Mukherjee, V. Pulabaigari, Hater-O-genius aggression classification using capsule networks, in: Proceedings of the 17th International Conference on Natural Language Processing (ICON), 2020.
[30] A. C. Mazari, N. Boudoukhani, A. Djeffal, Bert-based ensemble learning for multi-aspect hate speech detection, Cluster Computing (2023) 1–15.
[31] N. S. Samghabadi, P. Patwa, S. Pykl, P. Mukherjee, A. Das, T. Solorio, Aggression and misogyny detection using bert: A multi-task approach, in: Proceedings of the second workshop on trolling, aggression and cyberbullying, 2020.
[32] J. Risch, R. Krestel, Bagging BERT models for robust aggression identification, in: Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying, 2020.
[33] S. Nagar, F. A. Barbhuiya, K. Dey, Towards more robust hate speech detection: using social context and user data, Social Network Analysis and Mining 13 (2023) 47.
[34] A. Laddha, M. Hanoosh, D. Mukherjee, P. Patwa, A. Narang, Understanding chat messages for sticker recommendation in messaging apps, Proceedings of the AAAI Conference on Artificial Intelligence 34 (2020). doi:10.1609/aaai.v34i08.7019.
[35] A. Laddha, M. Hanoosh, D. Mukherjee, P. Patwa, A. Narang, Large scale multilingual sticker recommendation in messaging apps, AI Magazine 42 (2022). doi:10.1609/aaai.12023.
[36] P. Patwa, G. Aguilar, S. Kar, S. Pandey, S. PYKL, B. Gambäck, T. Chakraborty, T. Solorio, A. Das, SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets, in: Proceedings of the Fourteenth Workshop on Semantic Evaluation, 2020.
[37] B. R. Chakravarthi, et al., 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, 2021.
[38] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, J. P. McCrae, Corpus creation for sentiment analysis in code-mixed tamil-english text, arXiv:2006.00206 (2020).
[39] B. R. Chakravarthi, N. Jose, S. Suryawanshi, E. Sherly, J. P. McCrae, A sentiment analysis dataset for code-mixed malayalam-english, arXiv preprint arXiv:2006.00210 (2020).
[40] A. Hande, R. Priyadharshini, B. R. Chakravarthi, KanCMD: Kannada CodeMixed dataset for sentiment analysis and offensive language detection, in: Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media, 2020.
[41] S. Dowlagar, R. Mamidi, Graph convolutional networks with multi-headed attention for code-mixed sentiment analysis, in: Proceedings of the First Workshop on Speech and

Language Technologies for Dravidian Languages, 2021, pp. 65–72.
[42] J. Risch, A. Stoll, M. Ziegele, R. Krestel, hpidedis at germeval 2019: Offensive language identification using a german bert model., in: KONVENS, 2019.
[43] D. Tula, P. Potluri, S. Ms, S. Doddapaneni, P. Sahu, R. Sukumaran, P. Patwa, Bitions@DravidianLangTech-EACL2021: Ensemble of multilingual language models with pseudo labeling for offence detection in Dravidian languages, in: Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages, 2021.
[44] D. Tula, M. S. Shreyas, V. Reddy, P. Sahu, S. Doddapaneni, P. Potluri, R. Sukumaran, P. Patwa, Offence detection in dravidian languages using code-mixing index-based focal loss, SN Computer Science 3 (2022). doi:10.1007/s42979-022-01190-1.
[45] Y. Ma, L. Zhao, J. Hao, XLP at SemEval-2020 task 9: Cross-lingual models with focal loss for sentiment analysis of code-mixing language, in: Semeval, 2020.
[46] M. Ali, S. T. Kandukuri, S. Manduru, P. Patwa, A. Das, Pesto: Switching point based dynamic and relative positional encoding for code-mixed languages (student abstract), AAAI (2022). doi:10.1609/aaai.v36i11.21587.
[47] A. Hu, S. Flaxman, Multimodal sentiment analysis to explore the structure of emotions, in: KDD, 2018.
[48] R. Jha, V. Kaki, V. Kolla, S. Bhagat, P. Patwa, A. Das, S. Pal, Image2tweet: Datasets in Hindi and English for generating tweets from images, in: Proceedings of the 18th International Conference on Natural Language Processing (ICON), 2021.
[49] R. Gomez, J. Gibert, L. Gomez, D. Karatzas, Exploring hate speech detection in multimodal publications, 2019. arXiv:1910.03814.
[50] B. Zadeh, et al., Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph, in: ACL, 2018.
[51] D. Kiela, H. Firooz, A. Mohan, V. Goswami, A. Singh, P. Ringshia, D. Testuggine, The hateful memes challenge: Detecting hate speech in multimodal memes, Neurips (2020).
[52] S. Suryawanshi, B. R. Chakravarthi, Findings of the shared task on troll meme classification in Tamil, in: Speech and Language Technologies for Dravidian Languages, 2021.
[53] E. Hossain, O. Sharif, M. M. Hoque, MUTE: A multimodal dataset for detecting hateful memes, in: Proceedings of the 2nd AACL Student Research Workshop, 2022.
[54] Z. Xie, L. Liu, Y. Wu, L. Zhong, L. Li, Learning text-image joint embedding for efficient cross-modal retrieval with deep feature engineering, ACM Transactions on Information Systems 40 (2021) 1–27. URL: https://doi.org/10.1145%2F3490519. doi:10.1145/3490519.
[55] V. Krishna, S. Suryavardan, S. Mishra, S. Ramamoorthy, P. Patwa, M. Chakraborty, A. Chadha, A. Das, A. Sheth, Imaginator: Pre-trained image+text joint embeddings using word-level grounding of images, 2023. arXiv:2305.10438.
[56] N. Gunti, S. Ramamoorthy, P. Patwa, A. Das, Memotion analysis through the lens of joint embedding (student abstract), Proceedings of the AAAI Conference on Artificial Intelligence 36 (2022). doi:10.1609/aaai.v36i11.21616.
[57] J. Lu, D. Batra, D. Parikh, S. Lee, Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks, 2019. arXiv:1908.02265.
[58] L. H. Li, M. Yatskar, D. Yin, C.-J. Hsieh, K.-W. Chang, Visualbert: A simple and performant baseline for vision and language, 2019. arXiv:1908.03557.
[59] S. Ramamoorthy, N. Gunti, S. Mishra, S. Suryavardan, A. Reganti, P. Patwa, A. Das,

T. Chakraborty, A. Sheth, A. Ekbal, et al., Memotion 2: Dataset on sentiment and emotion analysis of memes, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, 2022.
[60] K. N. Phan, G.-S. Lee, H.-J. Yang, S.-H. Kim, Little flower at memotion 2.0 2022: Ensemble of multi-modal model using attention mechanism in memotion analysis, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, 2022.
[61] T. Morishita, G. Morio, S. Horiguchi, H. Ozaki, T. Miyoshi, Hitachi at SemEval-2020 task 8: Simple but effective modality ensemble for meme emotion recognition, in: SemEval, 2020.
[62] Y. Guo, J. Huang, Y. Dong, M. Xu, Guoym at SemEval-2020 task 8: Ensemble-based classification of visuo-lingual metaphor in memes, in: SemEval, 2020.
[63] G.-A. Vlad, G.-E. Zaharia, D.-C. Cercel, et al., Upb at semeval-2020 task 8: Joint textual and visual modeling in a multi-task learning architecture for memotion analysis, 2020. arXiv:2009.02779.
[64] T. T. Nguyen, N. T. Pham, N. D. Nguyen, et al., Hcilab at memotion 2.0 2022: Analysis of sentiment, emotion and intensity of emotion classes from meme images using single and multi modalities, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, 2022.
[65] G. G. Lee, M. Shen, Amazon pars at memotion 2.0 2022: Multi-modal multi-task learning for memotion 2.0 challenge, Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection (2020).
[66] M. Bhange, N. Kasliwal, Hinglishnlp: Fine-tuned language models for hinglish sentiment detection, arXiv preprint arXiv:2008.09820 (2020).
[67] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., An image is worth 16x16 words: Transformers for image recognition at scale, arXiv:2010.11929 (2020).
[68] W. Yu, D. Kolossa, wentaorub at Memotion 3: Ensemble learning for multi-modal meme classification, in: Proceedings of De-Factify 2: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2023.
[69] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, et al., Learning transferable visual models from natural language supervision, in: ICML, 2021.
[70] X. Li, X. Yin, C. Li, P. Zhang, et al., Oscar: Object-semantics aligned pre-training for vision-language tasks, 2020. arXiv:2004.06165.
[71] Y.-C. Tang, K.-D. Wang, T.-Y. Ou, W.-C. Peng, NYCU_TWO at Memotion 3: Good foundation, good teacher, then you have good meme analysis, in: Proceedings of De-Factify 2: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2023.
[72] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, 2021. arXiv:2103.14030.
[73] X. Guo, J. Ma, A. Zubiaga, NUAA-QMUL-AIIT at Memotion 3: Multi-modal fusion with squeeze-and-excitation for internet meme emotion analysis, in: Proceedings of De-Factify 2: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2023.
[74] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach (2019).
[75] X. Zhai, J. Puigcerver, A. Kolesnikov, et al., A large-scale study of representation learning with the visual task adaptation benchmark, 2020. arXiv:1910.04867.
[76] G. Ke, Q. Meng, T. Finley, et al., Lightgbm: A highly efficient gradient boosting decision

tree, in: Neurips, 2017.
[77] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: CVPR, 2016.
[78] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A unified embedding for face recognition and clustering, in: CVPR, 2015.