TeamX@DravidianLangTech-ACL2022: A Comparative Analysis for Troll-Based Meme Classification

Rabindra Nath Nandi[1], Firoj Alam[2], Preslav Nakov[2]
[1]BJIT Limited, Dhaka, Bangladesh
[2]Qatar Computing Research Institute, HBKU, Doha, Qatar
`rabindra.nath@bjitgroup.com`, {`falam,pnakov`}`@hbku.edu.qa`

Abstract

The spread of fake news, propaganda, misinformation, disinformation, and harmful content online raised concerns among social media platforms, government agencies, policymakers, and society as a whole. This is because such harmful or abusive content leads to several consequences to people such as physical, emotional, relational, and financial. Among different harmful content *trolling-based* online content is one of them, where the idea is to post a message that is provocative, offensive, or menacing with an intent to mislead the audience. The content can be textual, visual, a combination of both, or a meme. In this study, we provide a comparative analysis of troll-based memes classification using the textual, visual, and multimodal content. We report several interesting findings in terms of code-mixed text, multimodal setting, and combining an additional dataset, which shows improvements over the majority baseline.

1 Introduction

Social media have become one of the main communication channels for the propagation of information through textual, visual, or audio-visual content. While the content shared on social media creates a positive impact, however, there are also content that spread harm and hostility (Brooke, 2019), including abusive language (Mubarak et al., 2017), propaganda (Da San Martino et al., 2020, 2019) cyberbullying (Van Hee et al., 2015), cyber-aggression (Kumar et al., 2018), and other kinds of harmful content (Pramanick et al., 2021). The propagation of such content most often is done by an automated tool, troll, or coordinated groups, which target specific users, communities (e.g., minority groups), individuals, and companies. To detect such content there has been effort to develop automatic tools (see most recent surveys on disinformation (Alam et al., 2021c), rumours (Bondielli and Marcelloni, 2019), propaganda (Da San Martino et al., 2020),

and multimodal memes (Afridi et al., 2021), hate speech (Fortuna and Nunes, 2018), cyberbullying (Haidar et al., 2016), and offensive content (Husain and Uzuner, 2021). In addition, shared tasks has also been organized in the past years addressing factuality, fake news and harmful content (Nakov et al., 2021; Kiela et al., 2020).

Among other social media content, recently, the uses of *internet memes* became popular and they are often shared for the purpose of humor or fun with no bad intentions. However, memes are also created and shared with harmful intentions. This include attack on people based on the characteristics such as ethnicity, race, sex, gender identity, disability, disease, nationality, and immigration status (Kiela et al., 2020). There has been research effort to develop computational method to detect such memes, such as detecting hateful memes (Kiela et al., 2020), propaganda (Dimitrov et al., 2021a,b), offensive (Suryawanshi et al., 2020a), sexist meme (Fersini et al., 2019) and troll based meme (Suryawanshi and Chakravarthi, 2021).

In this study, we focus on troll-based meme classification based on the dataset released in the shared task discussed in (Suryawanshi et al., 2022). While meme contains both textual and visual elements, hence, we investigate textual, visual content and their combination using different pretraind transformer models. In addition, we explored combining an external dataset, and use of code-mixed text (i.e., Tamil and English) extracted using OCR. Note that the text provided with the dataset is transcribed in Latin. While prior work focuses on the text provided with the dataset, here, we also follow a different strategy, directly using the text from the OCR without any cleaning.

Our contributions include:

- we investigate classical algorithm (e.g., SVM), pretraind transformer and deep CNN models for both text and images, respectively;

- we combine an additional dataset and use code-mixed text, extracted using OCR and compare the performance;

- we also experiment with different pretrained multimodal models.

2 Related Work

Prior work on detecting harmful aspects of memes include categorizing hateful memes (Kiela et al., 2020), antisemitism (Chandra et al., 2021) and propaganda detection techniques in memes (Dimitrov et al., 2021a), harmful memes and their target (Pramanick et al., 2021), identifying protected category such as race, sex that has been attacked (Zia et al., 2021), and identifying offensive content (Suryawanshi et al., 2020a). Among the studies most notable effort that streamlined the research work include shared tasks such as "Hateful Memes Challenge" (Kiela et al., 2020), detection of persuasion techniques (Dimitrov et al., 2021b) and troll meme classification (Suryawanshi and Chakravarthi, 2021).

The work by Chandra et al. (2021) investigates antisemitism along with its types by addressing the tasks as binary and multi-class classification using pretrained transformers and CNN as modality-specific encoders along with various multimodal fusion strategies. Dimitrov et al. (2021a) developed a dataset with 22 propaganda techniques and investigates the different state-of-the-art pretrained models and demonstrate that joint vision-language models perform best. Pramanick et al. (2021) address two tasks such as detecting harmful memes and identifying the social entities they target and propose a multimodal model, which utilizes local and global information. Zia et al. (2021) goes one step further than a binary classification of hateful memes – more fine-grained categorization based on protected category (i.e., race, disability, religion, nationality, sex) and their attack types (i.e., contempt, mocking, inferiority, slurs, exclusion, dehumanizing, inciting violence) using the dataset released in the WOAH 2020 Shared Task.[1] Fersini et al. (2019) studied sexist meme detection and investigate textual cues with a late-fusion strategy, which suggest that fusion approach performs better. The same authors also developed a dataset of 800 misogynistic memes covering different manifestations of hatred against women (e.g., body shaming, stereotyping, objectification and violence), which are collected from different social media (Gasparini et al., 2021).

In the "Hateful Memes Challenge", the participants addressed the hateful meme classification task by fine-tuning the state-of-art multi-modal transformer models (Kiela et al., 2021) and best system in the competition used different unimodal and multimodal pre-training models such as VisualBERT (Li et al., 2019) VL-BERT (Su et al., 2019), UNITER (Chen et al., 2019), VILLA (Gan et al., 2020) and ensembles (Kiela et al., 2021). The SemEval-2021 propaganda detection shared task (Dimitrov et al., 2021b) was organized with a focus on fine-grained propaganda techniques in text and the entire meme, and from the participants' systems, they conclude that multimodal cues are important for automated propaganda detection. In the troll meme classification shared task (Suryawanshi and Chakravarthi, 2021), the best system used ResNet152, BERT with multimodal attention, and the majority of the system used pretrained transformer models for text, CNN models for images, and early fusion approaches.

3 Experiments

3.1 Data

We use the dataset provided in the troll-based Tamil meme classification shared task discussed in (Suryawanshi et al., 2020b, 2022). The dataset is comprised of meme and transcribed text in Latin, which are annotated and transcribed by native Tamil speakers. There is a total of 2,300, 667 memes for training and testing, respectively. For our experiments, we split the training set into training and development set with 80 and 20%, respectively. The development set is used for fine-tuning the models.

In Table 1, we report the distribution of the dataset that we used for the experiments.

Class	Train	Dev	Test
Troll	1013	269	395
Non-Troll	827	191	272

Table 1: Distribution of the Troll-based Tamil Meme dataset. We split original training set into training and development set.

---

[1]github.com/facebookresearch/fine_grained_hateful_memes

3.2 Settings

For the classification, we run different unimodal experiments: *(i)* only text, *(ii)* only meme, and *(iii)* text and meme together. For each setting, we also run several baseline experiments. One such baseline is the *majority class* baseline, which predicts the label based on the most frequent label in the training set. This has been most commonly used in shared tasks (Nakov et al., 2021). Furthermore, we run a few advanced experiments using an additional dataset and code-mixed text from OCR. To measure the performance of each model we used a weighted F_1 score to maintain shared task guideline.

3.2.1 Text Modality

For the baseline using text modality, we used bag-of-n-gram vectors weighted with logarithmic term frequencies (tf) multiplied with inverse document frequencies (idf) and train the model using Support Vector Machines (SVM) (Platt, 1998). Note that we extracted unigram, bigram, and tri-gram features. We used grid search to optimize the SVM hyperparameters.

We then experiment using multilingual BERT (mBERT) model (Devlin et al., 2019). We performed ten reruns for each experiment using different random seeds, then we picked the model that performed best on the development set. We used a batch size of 8, a learning rate of 2e-5, maximum sequence length 128, three epochs, and used the 'categorical cross-entropy' as the loss function.

3.2.2 Image Modality

Similar to the text modality, for the baseline experiment with image modality, we extract features from a pre-trained model, then train the model using SVM. We extracted features from the penultimate layer of the EfficientNet (b1) model (Tan and Le, 2019), which was trained using ImageNet. For training the model using SVM we used the default parameters setting.

For the later experiments we used the transfer learning approach, fine-tuning the pre-trained deep CNN models (e.g., VGG16), which has been shown success for visual recognition tasks. We used the weights of the model pre-trained on ImageNet to initialize our model. We adapt the last layer (i.e., softmax layer) of the network for the binary classification task. We trained models using three popular neural network architectures such as VGG16 (Simonyan and Zisserman, 2015), ResNet101 (He

et al., 2016) and EfficientNet (Tan and Le, 2019), which showed state-of-art performance in similar tasks (Ofli et al., 2020; Alam et al., 2021a,b). For training, we used the Adam optimizer (Kingma and Ba, 2015) with an initial learning rate of 10^{-5}, which is decreased by a factor of 10 when accuracy on the dev set stops improving for 10 epochs. From our experiment, we observe that the model converse within 50-60 epochs. As the dataset size is small, fine-tuning the entire network did not yield better results, therefore, we freeze the network and fine-tune the penultimate layer.

3.2.3 Multimodal: Text and Image

For the multimodel experiments, we used Vision Transformer (ViT) (Kolesnikov et al., 2021) for image feature encoding and multilingual BERT (mBERT) for the textual representation. We used BLOCK fusion (Ben-younes et al., 2019) to merge the features from two different modalities. BLOCK fusion is a multimodal fusion based on block-superdiagonal tensor decomposition. Previously, Kolesnikov et al. (2021) showed a better performance using BLOCK fusion over several bilinear fusion techniques for Visual Question Answering (VQA) and Visual Relationship Detection (VRD) tasks. We conduct two major experiments by varying the textual data, *(i)* using the provided text, and *(ii)* using the code-mixed text.

3.2.4 Additional Experiments

Code-mixed Tamil and English text: The dataset comes with extracted text in a transcribed form in Latin. Given that current multilingual transformer models have not trained using such a latin form of text, therefore, to further understand the problem we extract text from memes using *tesseract* (Smith, 2007).[2] We then train the same mBERT model using the extracted code-mixed Tamil and English text. Note that the extracted text contains noise that comes from the output of the OCR.

Additional data: While we visually inspected the dataset we realized that textual and visual elements of memes have similarities with memes in hateful meme dataset (Kiela et al., 2020). We then mapped the labels from hateful memes dataset to troll labels, *(i)* hateful to *troll*, and *(ii)* not-hateful to *non-troll*. We combined all training and test set memes from the hateful memes dataset with the training set of the troll meme dataset. The dev set

---

[2]https://pypi.org/project/pytesseract/

Exp	Acc	P	R	F1
Maj.	59.2	35.1	59.2	44.1
Text modality				
Tf-Idf + SVM	46.2	47.9	46.2	**46.6**
mBERT	52.6	51.0	52.6	**51.4**
Add. data + mBERT	56.5	52.8	56.5	**51.8**
Code-mixed text + mBERT	57.3	55.2	57.3	**55.2**
Image modality				
EffNet feat +SVM	59.1	55.6	59.1	**50.1**
VGG16	55.8	50.2	55.8	**49.2**
ResNet101	60.4	58.6	60.4	**53.8**
EffNet (b1)	61.5	61.5	61.5	**53.6**
Multimodality				
ViT + mBERT	57.9	55.1	57.9	**54.2**
ViT + mBERT (code-mixed text)	55.3	57.0	55.3	**55.7**
Image modality + Additional data				
VGG16	54.4	52.0	54.4	**52.3**
ResNet101	60.7	58.9	60.7	**56.2**
EffNet (b1)	60.7	58.9	60.7	<u>**56.6**</u>

Table 2: Evaluation results on the test set. The results that improve over the majority class baseline are in **bold**, and the best one is <u>underlined</u>. Maj.: Majority baseline, Add. data: Additional data.

of the hateful meme dataset is combined with the dev set of the troll dataset. After combining the datasets we experiment with both text and image modalities using the same models.

4 Results and Discussion

In Table 2, we present the results for different modalities and settings. Overall, all results are better than majority baseline. In the unimodal experiments, the image-only models perform better than the text-only models. Our experiments on text-only models suggest that code-mixed data help in improving the performance using the same multiligual mBERT model compared to the transcribed latin text, which suggest that multilingual model is not able to capture the information in latin Tamil text. The additional data, which is in English only, slightly improved the performance for text-only experiment.

For the image-only experiments, we obtain a comparative performance with ResNet101 and EfficientNet (EffNet (b1)).

Our experiments on multimodel experiments provide better results compared to text and image modalities as shown in the Table 2. For the two

multimodel experiments, we obtained weighted-F1 scores of 54.2 and 55.7 for text and for code-mixed text, respectively, which are better than the best weighted-F1 score from text modality (51.4) and from image modality (53.8). Note that, the performance of our multimodel experiments is better than Hegde et al. (2021), where a weighted F1 score of 0.47 has been reported for a similar task. That implies BLOCK fusion has advantages over late fusion for troll meme classification tasks.

With additional data, results improved significantly for all image-only models. It confirms our observation that visual elements in memes share information across different datasets, and possibly different cultures too.

Our future plan is to apply the data augmentation technique for multimodality experiments and also we have a plan to explore the performance of popular multimodel algorithms.

During the shared task participation, we submitted one run, using only image modality, where our system was not showing promising performance. However, our subsequent experiments and analysis show several promising directions in terms of original code-mixed data and additional data from the other task.

5 Conclusion and Future Work

We present a comparative analysis of different modalities for the troll-based meme classification task. We show that the multilingual model capture more information for code-mixed text than its Latin counterpart. We present higher performance with multimodality compared to unimodal models. Our experiments also suggest that additional data from the other task helps in capturing visual information. In the future, we plan to further develop multimodal models that can capture information in code-mixed noisy conditions.

References

Tariq Habib Afridi, Aftab Alam, Muhammad Numan Khan, Jawad Khan, and Young Koo Lee. 2021. A multimodal memes classification: A survey and open research issues. In *5th International Conference on Smart City Applications, SCA 2020*, pages 1451–1466. Springer Science and Business Media Deutschland GmbH.

Firoj Alam, Tanvirul Alam, Md Hasan, Abul Hasnat, Muhammad Imran, Ferda Ofli, et al. 2021a. MEDIC: a multi-task learning dataset for disaster image classification. *arXiv:2108.12828*.

Firoj Alam, Tanvirul Alam, Muhammad Imran, and Ferda Ofli. 2021b. Robust training of social media image classification models for rapid disaster response. *arXiv:2104.04184*.

Firoj Alam, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimiter Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, and Preslav Nakov. 2021c. A survey on multimodal disinformation detection. *arXiv:2103.12541*.

Hedi Ben-younes, Remi Cadene, Nicolas Thome, and Matthieu Cord. 2019. Block:bilinear superdiagonal fusion for visual question answeringand visual relationship detection. In *Proceedings of the 33st Conference on Artificial Intelligence (AAAI)*, volume 15.

Alessandro Bondielli and Francesco Marcelloni. 2019. A survey on fake news and rumour detection techniques. *Information Sciences*, 497:38–55.

Sian Brooke. 2019. "Condescending, Rude, Assholes": Framing gender and hostility on Stack Overflow. In *Proceedings of the Third Workshop on Abusive Language Online*, pages 172–180, Florence, Italy. Association for Computational Linguistics.

Mohit Chandra, Dheeraj Pailla, Himanshu Bhatia, Aadilmehdi Sanchawala, Manish Gupta, Manish Shrivastava, and Ponnurangam Kumaraguru. 2021. "Subverting the Jewtocracy": Online antisemitism detection using multimodal deep learning. *arXiv:2104.05947*.

Yen-Chun Chen, Linjie Li, Licheng Yu, Ahmed El Kholy, Faisal Ahmed, Zhe Gan, Yu Cheng, and Jingjing Liu. 2019. UNITER: Learning universal image-text representations.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020. A survey on computational propaganda detection. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 4826–4832. ijcai.org.

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5636–5646, Hong Kong, China. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021a. Detecting propaganda techniques in memes. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, ACL-IJCNLP '21, pages 6603–6617, Online. Association for Computational Linguistics.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021b. Task 6 at SemEval-2021: Detection of persuasion techniques in texts and images. In *Proceedings of the 15th International Workshop on Semantic Evaluation*, SemEval '21, Bangkok, Thailand. Association for Computational Linguistics.

Elisabetta Fersini, Francesca Gasparini, and Silvia Corchs. 2019. Detecting sexist meme on the web: A study on textual and visual cues. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pages 226–231. IEEE.

Paula Fortuna and Sérgio Nunes. 2018. A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, 51(4):1–30.

Zhe Gan, Yen-Chun Chen, Linjie Li, Chen Zhu, Yu Cheng, and Jingjing Liu. 2020. Large-scale adversarial training for vision-and-language representation learning. *Advances in Neural Information Processing Systems*, 33:6616–6628.

Francesca Gasparini, Giulia Rizzi, Aurora Saibene, and Elisabetta Fersini. 2021. Benchmark dataset of memes with text transcriptions for automatic detection of multi-modal misogynistic content. *CoRR*, abs/2106.08409.

Batoul Haidar, Maroun Chamoun, and Fadi Yamout. 2016. Cyberbullying detection: A survey on multilingual techniques. In *2016 European Modelling Symposium (EMS)*, pages 165–171.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Siddhanth U Hegde, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. Uvce-iiitt@ dravidianlangtech-eacl2021: Tamil troll meme classification: You need to pay more attention. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 180–186.

Fatemah Husain and Ozlem Uzuner. 2021. A survey of offensive language detection for the arabic language. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 20(1):1–44.

Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Casey A Fitzpatrick, Peter Bull, Greg Lipstein, Tony Nelli, Ron Zhu, et al. 2021. The hateful memes challenge: competition report. In *NeurIPS 2020 Competition and Demonstration Track*, pages 344–360. PMLR.

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. In *Advances in Neural Information Processing Systems*, Online.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.

Alexander Kolesnikov, Alexey Dosovitskiy, Dirk Weissenborn, Georg Heigold, Jakob Uszkoreit, Lucas Beyer, Matthias Minderer, Mostafa Dehghani, Neil Houlsby, Sylvain Gelly, Thomas Unterthiner, and Xiaohua Zhai. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*.

Ritesh Kumar, Atul Kr Ojha, Shervin Malmasi, and Marcos Zampieri. 2018. Benchmarking aggression identification in social media. In *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying*, pages 1–11.

Liunian Harold Li, Mark Yatskar, Da Yin, Cho-Jui Hsieh, and Kai-Wei Chang. 2019. VisualBERT: A simple and performant baseline for vision and language. *arXiv:1908.03557*.

Hamdy Mubarak, Kareem Darwish, and Walid Magdy. 2017. Abusive language detection on arabic social media. In *Proceedings of the first workshop on abusive language online*, pages 52–56.

Preslav Nakov, Da San Martino Giovanni, Tamer Elsayed, Alberto Barrón-Cedeño, Rubén Míguez, Shaden Shaar, Firoj Alam, Fatima Haouari, Maram Hasanain, Watheq Mansour, Bayan Hamdan, Zien Sheikh Ali, Nikolay Babulkov, Alex Nikolov, Gautam Kishore Shahi, Julia Maria Struß, Thomas Mandl, Mucahid Kutlu, and Yavuz Selim Kartal. 2021. Overview of the CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news. LNCS (12880). Springer.

Ferda Ofli, Firoj Alam, and Muhammad Imran. 2020. Analysis of social media data using multimodal deep learning for disaster response. In *Proceedings of the Information Systems for Crisis Response and Management*.

John Platt. 1998. Sequential minimal optimization: A fast algorithm for training support vector machines.

Shraman Pramanick, Shivam Sharma, Dimitar Dimitrov, Md. Shad Akhtar, Preslav Nakov, and Tanmoy Chakraborty. 2021. MOMENTA: A multimodal framework for detecting harmful memes and their targets. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4439–4455, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations*, an Diego, CA, USA.

Ray Smith. 2007. An overview of the tesseract ocr engine. In *Ninth international conference on document analysis and recognition*, volume 2, pages 629–633.

Weijie Su, Xizhou Zhu, Yue Cao, Bin Li, Lewei Lu, Furu Wei, and Jifeng Dai. 2019. VL-BERT: Pretraining of generic visual-linguistic representations. *arXiv:1908.08530*.

Shardul Suryawanshi and Bharathi Raja Chakravarthi. 2021. Findings of the shared task on Troll Meme Classification in Tamil. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Mihael Arcan, and Paul Buitelaar. 2020a. 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).

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Mihael Arcan, Susan Levy, Paul Buitaleer, Prasanna Kumar Kumaresan, Rahul Ponnusamy, and Adeep Hande. 2022. Findings of the second shared task on Troll Meme Classification in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Pranav Verma, Mihael Arcan, John Philip McCrae, and Paul Buitelaar. 2020b. A dataset for troll classification of TamilMemes. In *Proceedings of the WILDRE5– 5th Workshop on Indian Language Data: Resources and Evaluation*, pages 7–13, Marseille, France. European Language Resources Association (ELRA).

Mingxing Tan and Quoc V Le. 2019. EfficientNet: rethinking model scaling for convolutional neural networks. *arXiv:1905.11946*.

Cynthia Van Hee, Els Lefever, Ben Verhoeven, Julie Mennes, Bart Desmet, Guy De Pauw, Walter Daelemans, and Veronique Hoste. 2015. Detection and fine-grained classification of cyberbullying events. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 672–680, Hissar, Bulgaria. INCOMA Ltd. Shoumen, BULGARIA.

Haris Bin Zia, Ignacio Castro, and Gareth Tyson. 2021. Racist or sexist meme? classifying memes beyond hateful. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pages 215–219, Online. Association for Computational Linguistics.