



TotalDefMeme: A Multi-Attribute Meme dataset on Total Defence in Singapore

Nirmalendu Prakash*

Singapore University of
Technology and Design
Singapore, Singapore

nirmalendu_prakash@sutd.edu.sg

Ming Shan Hee*

Singapore University of
Technology and Design
Singapore, Singapore

mingshan_hee@mymail.sutd.edu.sg

Roy Ka-Wei Lee

Singapore University of
Technology and Design
Singapore, Singapore

roy_lee@sutd.edu.sg

ABSTRACT

Total Defence is a defence policy combining and extending the concept of military defence and civil defence. While several countries have adopted total defence as their defence policy, very few studies have investigated its effectiveness. With the rapid proliferation of social media and digitalisation, many social studies have been focused on investigating policy effectiveness through specially curated surveys and questionnaires either through digital media or traditional forms. However, such references may not truly reflect the underlying sentiments about the target policies or initiatives of interest. People are more likely to express their sentiment using communication mediums such as starting topic thread on forums or sharing memes on social media. Using Singapore as a case reference, this study aims to address this research gap by proposing TotalDefMeme, a large-scale multi-modal and multi-attribute meme dataset that captures public sentiments toward Singapore's Total Defence policy. Besides supporting social informatics and public policy analysis of the Total Defence policy, TotalDefMeme can also support many downstream multi-modal machine learning tasks, such as aspect-based stance classification and multi-modal meme clustering. We perform baseline machine learning experiments on TotalDefMeme and evaluate its technical validity, and present possible future interdisciplinary research directions and application scenarios using the dataset as a baseline.

CCS CONCEPTS

• Computing methodologies → Natural language processing; Computer vision representations.

KEYWORDS

multimodal, meme, dataset, topic clustering, stance classification

ACM Reference Format:

Nirmalendu Prakash, Ming Shan Hee, and Roy Ka-Wei Lee. 2023. TotalDefMeme: A Multi-Attribute Meme dataset on Total Defence in Singapore. In *Proceedings of the 14th ACM Multimedia Systems Conference (MMSys '23)*.

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions.acm.org.

MMSys '23, June 7–10, 2023, Vancouver, BC, Canada

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0148-1/23/06...\$15.00

<https://doi.org/10.1145/3587819.3592545>

June 7–10, 2023, Vancouver, BC, Canada. ACM, New York, NY, USA, 7 pages.
<https://doi.org/10.1145/3587819.3592545>

1 INTRODUCTION

The growing popularity of social media platforms has led to the popularisation of internet memes, which provide the means for self-expression, connection, and social influence. Memes, typically presented as images with short superimposed text, are often designed to be relatable, humorous, and shareable. While memes have slowly become a fundamental aspect of contemporary digital culture, where most of the memes are humorous and meant for comic relief, there are increasingly more that are meant to express grievances and negative sentiments towards public individuals and institutions. Soh [19] highlights memes as a strategy for pragmatic resistance, allowing users to evade charges of sedition due to the authorship dispersal nature of memes. This meant that it was possible to disseminate information under the masquerade of humour, and craft narratives to shape public sentiments without much consideration of repercussions. Therefore, memes are a good communication medium for investigating the public's sentiment toward a country's government and its policies.

An example of such a policy may be encompassed under a country's defence strategy. *Total Defence* is a defence policy combining and extending the concept of military defence and civil defence. Several countries, including Sweden, Switzerland, Russia, Ukraine, Singapore, etc., have adopted the policy framework. It entails a high level of readiness of the state and its society to defend itself in catastrophic events such as war, crisis, and natural disasters. For instance, Singapore's *Total Defence* strategy focuses on six pillars in preparing and assessing the countries' readiness for catastrophic events: *military, civil, economic, social, psychological, and digital*. While several countries have adopted the *Total Defence* framework as a policy guidance, very few studies have investigated its effectiveness through the ground perceptions of the masses apart from specially curated surveys which may not reflect the underlying sentiments of society. This is primarily due to the unavailability of datasets to facilitate such an analysis.

This study aims to fill the data limitation gap by constructing TotalDefMeme^{1 2}, a large-scale multimodal and multi-attribute meme dataset that captures public sentiments toward Singapore's total defence policy. Figure 1 illustrates an example meme in our TotalDefMeme dataset. The meme depicts paramedics doing a good

¹https://gitlab.com/bottle_shop/meme/TotalDefMemes

²This dataset should only be used for non-commercial research purposes. Should you find any meme in violation of copyright, please notify us and we will take it down.



Figure 1: An example of a TotalDefMeme meme and its multi-attribute annotation. The meme praises Singapore’s Civil Defence by showing the paramedics going beyond their work duties to help an elderly woman.

deed beyond their work duties, expressing stances that are supportive of Singapore’s civil defence, which is of the country’s six total defence pillars.

Besides supporting social informatics and public policy analysis of total defence policy, TotalDefMeme can also support many downstream multimodal machine learning tasks. Existing meme research studies have extensively studied memes with malicious intent by analyzing and releasing large datasets on a few binary attributes, such as hate speech [8, 12], harmfulness [14], misinformation [15], offensiveness [20], information diffusion [10] etc. However, there has been limited focus on understanding the topics and stances expressed in memes. Furthermore, most of the existing meme datasets focused on the English language and western cultural context. In contrast, our TotalDefMeme is a multimodal meme dataset based on Southeast Asia and Singapore context, where the language captured is predominantly colloquial with a contextualized variant of English, otherwise known as Singlish³. Our TotalDefMeme dataset also aims to address the existing meme research gaps by providing a dataset with multiple attributes that could facilitate downstream machine learning tasks such as multimodal aspect-based stance analysis and multimodal meme clustering.

We summarize our contributions as follows:

- We construct TotalDefMeme, a large-scale multimodal and multi-attribute meme dataset that captures public sentiments toward Singapore’s total defence policy.
- We discuss and suggest possible future interdisciplinary research directions and application scenarios using the dataset.
- We perform a set of baseline machine learning experiments on TotalDefMeme and evaluate its technical validity.

³<https://en.wikipedia.org/wiki/Singlish>

2 RELATED WORKS

Meme Analysis. Multimodal study of memes has gained traction in recent years, which has resulted in multiple datasets with different objectives. Recently Facebook released a Hateful Memes Challenge[8], where the task is to classify memes as hateful and non-hateful. Mathias et al.[12] further extended the Hateful Memes Challenge dataset by adding the protected category (i.e., *race, disability, religion, nationality, sex*) that has been attacked by the meme, and the type of attack (i.e., *contempt, mocking, inferiority, slurs, exclusion, dehumanizing, inciting violence*). These hateful meme datasets have encouraged researchers to develop hateful meme detection solutions [1, 6, 9, 23]. Suryavanshi et al.[20] constructed a meme dataset on the 2016 United States presidential elections with offensiveness labels. Similarly, Pramanik et al. [14] released a dataset called *HarMeme*, to study harmful memes. The researchers retrieved memes by querying Google Image Search with keywords related to covid-19. As part of SemEval 2020 tasks, Sharma et al.[17] released a dataset called *Memotion*, to capture fine-grained emotions expressed in memes. Pramanick et al. [13] further proposed a multi-hop attention-based deep learning approach to leverage spatial-domain correspondence between visual and textual modalities to extract fine-grained feature representations for sentiment classification on the *Memotion* dataset. In a recent study, Qu et al.[15] examined memes that spread misinformation. They collect memes from Reddit on three topics, namely, *Covid-19, BLM*, and *veganism*, and annotate binary misinformation labels. While more meme datasets are available to support downstream multimodal machine learning tasks, most are annotated with simple binary labels and can only support limited classification tasks. Furthermore, most existing memes datasets only contain English memes based on western cultures.

Our study value-adds to the existing meme studies by providing a multimodal meme dataset that captures multiple attributes to support more multimodal machine learning tasks. Specifically, our proposed TotalDefMeme dataset contains memes annotated with multi-attribute labels, such as the types of memes, the topic discussed, the total defence pillars affected, and the stance towards the pillars. Furthermore, TotalDefMeme contains memes on Southeast Asia and Singapore’s cultural context, where the language captured is *Singlish*. The multi-attribute nature of our TotalDefMeme dataset also facilitates machine learning tasks, such as aspect-based stance analysis and multimodal topic clustering, that existing datasets could not support.

Total Defence. The Total Defence policy is a defence policy that incorporates both military and civil defence strategies. It is adopted by countries such as Finland, Norway, Sweden, and Singapore. The policy emphasizes a high level of readiness against any potential dangers and catastrophes, including war, crisis, and natural disasters, for the state and its society. Hence, as a key defence policy, Total Defence serves as the primary guide to the needs of the government and its citizens.

While various countries employ the Total Defence policy, their implementation is not globally uniform. Our research involved collecting and analyzing memes pertaining to Singapore’s Total Defence policy, which comprises six pillars: *Military, Civil, Economic, Social, Psychological*, and *Digital*. Through our examination

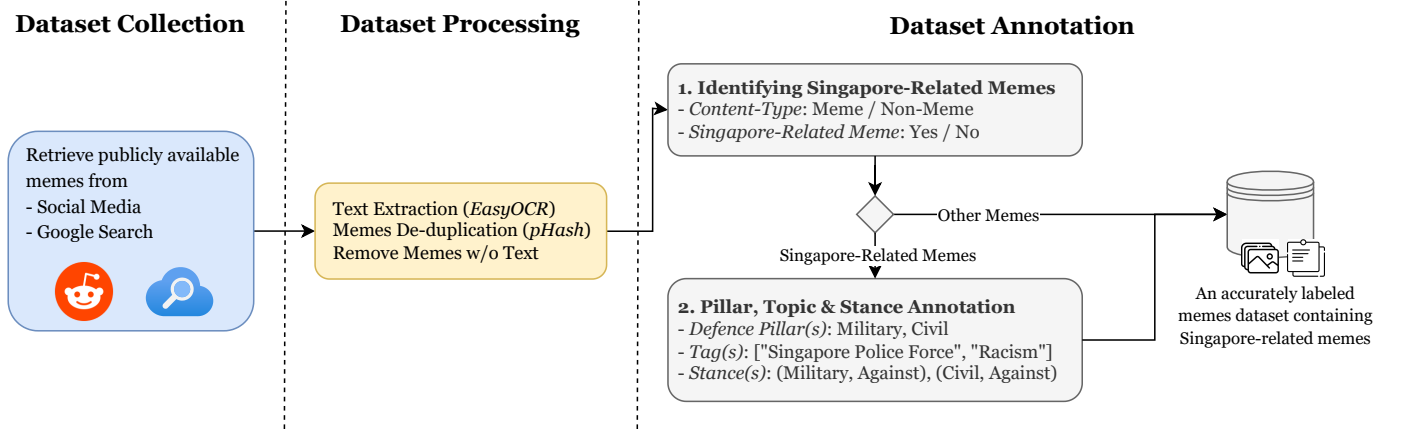


Figure 2: Data Collection & Annotation Pipeline.

of multiple sources⁴⁵, we have summarized the definitions of each pillar as follows:

- **Military Defence:** Strong and formidable defence force made up of Regulars and National Servicemen, and supported by the entire nation
- **Civil Defence:** Collective effort of the society to spot signs of threats, respond effectively and recover quickly from crisis
- **Economic Defence:** Strong and resilient economy that is globally competitive and able to bounce back from any crisis.
- **Social Defence:** Bonds that unite us, built on trust and understanding among people of different races and religions, living in harmony and looking out for one another
- **Psychological Defence:** The will and resolve to defend our way of life and interests, the fighting spirit to overcome challenges together
- **Digital Defence:** Being secure, alert and responsible online.

3 DATASET CONSTRUCTION

In this section, we describe the data annotation pipeline used for constructing the TotalDefMeme dataset and provide a preliminary analysis of the dataset. The pipeline, shown in Figure 2, mainly comprises three phases: *dataset collection*, *dataset processing*, and *dataset annotation*.

3.1 Dataset Collection

To collect memes related to Singapore’s Total Defence, we adopted a keyword-based approach using Google Search to obtain more relevant memes. We studied the Total Defence concepts across multiple sources and crafted the appropriate keywords such as “*police force*”, “*racism*”, “*phone scams*”. Subsequently, these keywords are inserted into a template query: “Singapore <keyword> Memes”. We further scraped various publicly available groups on popular social media platforms, such as Reddit and Instagram, to increase our coverage⁶. Including memes from social media platforms enables

Figure 3: Snapshot of the user interface used for *Meme Identification and Pillars, Topics & Stances annotation*

our examination of recent and viral Singapore-related memes. We obtained a dataset of 7,200 diverse memes through these methods.

3.2 Dataset Processing

To align with our research on multimodal memes, we applied strict filtering criteria. First, we performed a simple filtering on the quality of the memes where we removed memes with image resolution smaller than 224x224 pixels and memes with text exceeding 50-word tokens. Second, we applied a text extraction tool using the EasyOCR⁷ algorithm to retrieve the text found within the meme and remove the ones without text. Third, we utilized the pHash algorithm to identify groups of duplicates and preserved each group’s memes with the highest resolution. Finally, we retain 5,301 memes, which serves as the final TotalDefMeme dataset.

3.3 Dataset Annotation

We recruited six annotators who are familiar with Singapore’s culture and knowledgeable about Singapore’s Total Defence concept from a pool of shortlisted Singaporean undergraduates who have

⁴https://www.mindef.gov.sg/oms/imindef/mindef_websites/topics/totaldefence/about.html

⁵<https://www.scdf.gov.sg/home/community-volunteers/community-preparedness/total-defence>

⁶Note that we did not use content from any private or restricted pages

⁷<https://www.jaied.ai/easyocr/>

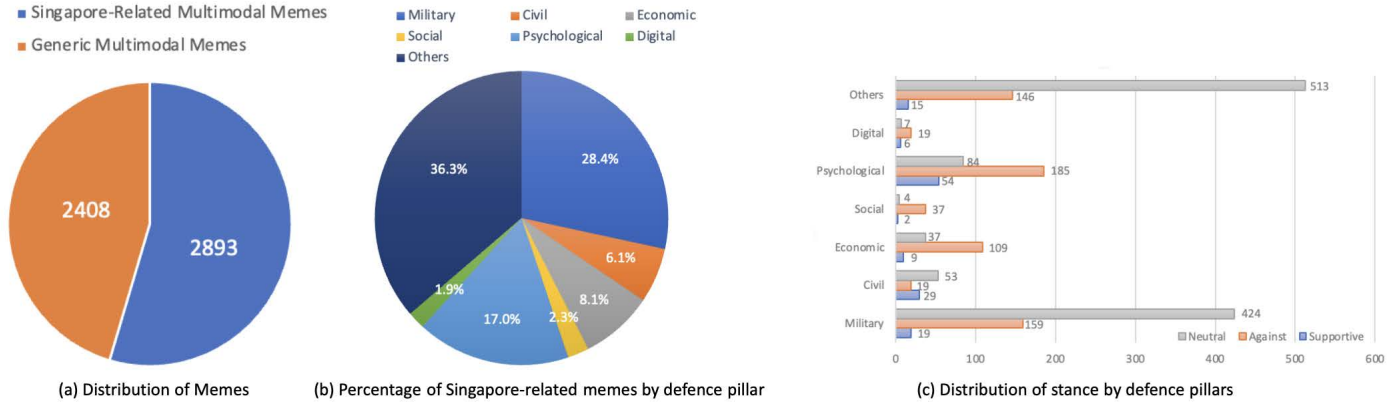


Figure 4: Statistics about TotalDefMeme dataset

Table 1: Krippendorff’s Alpha for each task: Meme Type (3 classes), Defence Pillar (7 classes) and Stance (3 classes).

Types	Pillars	Stances
0.65	0.55	0.21

passed a short screening survey. The annotators are tasked with identifying the Singapore-related memes and annotating the pillars, topics, and stances of these Singapore-related memes. Screenshots of the user interface can be seen in Figure 3

Meme Region Identification The annotators will conduct a manual review and distinguish Singapore-related multi-modal memes from the diverse collection of visuals, including infographics and posters. Each meme can be categorised into two types: generic memes and Singapore-related memes. For example, a meme that addresses the topic of working overtime shall be considered a generic meme, as this scenario can occur in the majority of countries, while a meme that addresses the Singapore Police Force shall be considered a Singapore-related meme.

Pillars, Topics & Stances Annotation Upon identifying the Singapore-related meme, the annotators annotate the pillars, topics, and stances of the meme. The annotators will first assign the memes’ defence pillars: *military*, *civil*, *economic*, *social*, *psychological*, *digital*, or *others*. We included the “*others*” pillar as an option as there are memes that talk about daily occurrences and modern affairs, which do not involve any of the six defence pillars. Next, they annotate the relevant topic tags associated with the meme (i.e., nouns, pronouns, and phrases) in a free-text format. The annotators can enter the most appropriate tags and are encouraged to enter as many relevant tags as possible. Lastly, the annotators annotate the meme’s stances towards the assigned pillars: *support*, *against*, or *neutral*.

Quality Control Measures To ensure the reliability of the dataset, each meme is annotated by two annotators. If the disagreements contain similar opinions, the overlap annotations will be considered correct labels. However, if there are disagreements with entirely different perspectives, a third annotator will be brought in to provide an additional annotation for the meme. The overlapping annotations between at least two annotators will then be considered the correct labels. In the extreme case where all three annotators have different opinions, the meme will be flagged and removed

from the dataset. Finally, we held review discussions with the annotators to discuss the annotations with disagreements, allowing our annotators to receive feedback and improve their annotations.

3.4 Dataset Analysis

Figure 4 shows the statistical summary of the TotalDefMeme dataset. More than half of the memes in our dataset are Singapore-related. On the defence pillar annotation, we notice the class label imbalance. *Military* is the dominant class, with close to a third of the memes (32%) labeled to be related to military defence. Conversely, only 2.2% and 2.3% of the memes are labeled as *Digital* and *Social* classes, respectively. Interestingly, we also noted a substantial number of memes annotated as *Others*, which captures other Singapore-related but non total defence related topics, adding diversity to the TotalDefMeme dataset.

We computed Krippendorff’s Alpha score to measure the inter-annotator agreement for the various annotation labels. Table 1 presents the scores for memes’ *types*, *pillars*, and *stances*. Although the task requires in-depth contextual knowledge of Singapore and total defence, the annotators have achieved a moderate agreement with alpha scores of 0.65 and 0.55 for *types* and *pillars*, respectively, indicating quality annotations. The agreement for *stances* is weaker (alpha score of 0.21) due to the subjectivity of the task. Nevertheless, we will release the annotations and annotators’ id in our released dataset to facilitate further analysis.

4 INTERDISCIPLINARY RESEARCH AND APPLICATIONS

The purpose of publishing this unique dataset is to motivate social science and computer science researchers to explore interdisciplinary research and propose novel machine learning tasks. As a starting point, we foresee this dataset to have several applications and usage scenarios. A few examples are as follows:

- Analysis and assessment of country’s total defence readiness
- Multimodal aspect-based stance classification
- Multimodal meme clustering
- Domain adaption of meme analysis

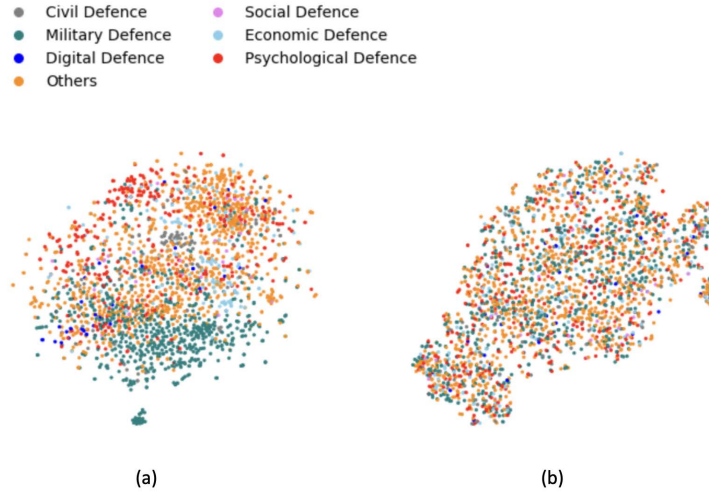


Figure 5: t-SNE visualization of (a) CLIP and (b) VisualBERT embeddings.

A clear use case for the TotalDefMeme dataset is to support interdisciplinary computational social science research. Computational social scientists can analyze the dataset to answer social informatics and policy-related questions. Furthermore, the annotation framework can also be applied to annotate and analyze total defence memes from other countries, such as Russia, Ukraine, etc.

Existing studies have explored sentiment and emotion classification in memes [17]. Our proposed dataset extends this line of research to support multimodal aspect-based stance classification, where the task aims to predict the stance towards an entity, concept, or events illustrated in a meme. Specifically, in TotalDefMeme, a possible task is to predict the stance towards the total defence pillar illustrated in a given meme. This is challenging as the machine learning model will need to interpret multi-modality information in the meme to predict both the total defence pillars and the corresponding stances. We will benchmark several baselines on this task using TotalDefMeme next Section 5.

Existing studies have explored performing clustering of memes using unimodal approaches [2, 4, 5, 7, 21, 22], where the memes' visuals are first captioned with text before applying topic models to cluster the memes or perform clustering on the memes' visuals representations. However, such approaches neglect the rich information in the interaction between textual and visual modality. Furthermore, evaluating the clustering algorithms is challenging as the ground truth is often unavailable. We hope that TotalDefMeme can encourage researchers to propose multimodal meme clustering methods. TotalDefMeme memes are also annotated with the topics and total defence pillars, which could serve as ground truth for evaluating the multimodal meme clustering algorithms. We will also benchmark several baselines on this task using TotalDefMeme.

Another potential use case of TotalDefMeme is to support domain adaption of meme analysis. Multimodal meme analysis techniques can be trained using TotalDefMeme and evaluate its performance on other meme datasets. This is particularly useful for analyzing total defence memes collected from other countries.

Table 2: Silhouette Score & NMI of the k-Mean clusters using various embeddings.

Embedding	Silhouette Score	NMI
BERT	0.030	0.020
VGG16	-0.013	0.010
VisualBERT	0.094	0.006
CLIP	0.012	0.091

Table 3: Accuracy scores for pillar and stance classification tasks.

Models	Pillar	Stance
BERT	0.18	0.30
VGG	0.14	0.30
VisualBERT	0.05	0.22
CLIP	0.57	0.54

5 EXPERIMENTS

5.1 Meme Clustering

To perform the clustering of memes, we first compute the memes' representation using various unimodal and multimodal models. Specifically, for unimodal models, we adopted pre-trained BERT [3] to compute a meme's representation using its texts, and VGG16 [18] to extract a meme's representation using its visual. For multimodal models, we use VisualBERT [11] to compute the meme's representation using both its visual and text. We also utilize CLIP [16] to extract a meme's visual and text representations separately before concatenating them to get the final meme's representation.

The computed meme's representations are subsequently input into k-mean clustering algorithm. We set the cluster size to 7, aiming to obtain pillar-wise semantic clusters. Table 2 shows the evaluation of the k-mean clustering results using the various unimodal and multimodal embeddings. Specifically, we computed the Silhouette and Normalized Mutual Information(NMI) scores for learned clusters. We observe that using the multimodal embeddings have achieved better clustering performance; VisualBERT embeddings has the highest Silhouette score (0.094), while CLIP embeddings





Meme	Nobody: NSF that went to Ex Crescendo waiting for their ORD blood test result: 	PM Lee to address the nation at 5PM NTUC workers: 	when u robbed a bank but forgot to turn off ur tracetogether bluetooth: 	NDP song you shouldn't sing during Hungry Ghost Month 
	(a)	(b)	(c)	(d)
Predicted Pillar(s)	Military	Psychological	Digital	Psychological
Ground-Truth Pillar(s)	Military	Economic	Digital	Psychological
Predicted Stance(s)	Military - Neutral	Psychological - Against	Digital - Against	Psychological - Against
Ground-Truth Stance(s)	Military - Neutral	Economic - Against	Digital - Against	Psychological - Neutral

Table 4: Example of CLIP pillars and stances classification results. Correct and incorrect predictions are color coded as green and red, respectively.

achieved the highest NMI (0.091). Conversely, using only the text (i.e., BERT) or visual (i.e., VGG16) embeddings results in poor clustering performance.

To further examine the multimodal embeddings, we show the t-SNE visualization of the k-mean clustering algorithm that was trained using CLIP and VisualBERT embeddings in Figure 5. The scatterplots are color-coded with the annotated total defence pillars. Interestingly, we observed that the CLIP embeddings seem to be clustering together according to the total defence pillar. In contrast, the VisualBERT embeddings seem to be scattered with no clear patterns based on pillars. Nevertheless, we note that the Silhouette score and NMI of simple k-mean clustering with multimodal embeddings are still very low, highlighting the challenge of this task and the need to develop better multimodal meme clustering methods.

5.2 Aspect-based Stance Classification

To perform aspect-based stance classification, we first stratified split TotalDefMeme by pillar into train, validation, and test sets using the ratio 60/20/20. Next, we train the baselines using the train set. For unimodal baselines, we fine-tune BERT [3] using the memes' text, and train VGG16 [18] using the memes' visual. For multimodal baselines, we train VisualBERT [11] and fine-tune CLIP [16] using both the memes' text and visual. As TotalDefMeme is an imbalance dataset, we upsampled the minority classes in the training set when training the model.

For this experiment, we formulate the aspect-based stance classification as a multitask classification problem where the model predicts both the pillars and pillars' stances as separate outputs. To achieve this, we modified the loss function in the linear output layer of the baseline classifiers to predict the memes' pillars and corresponding pillars' stances. Table 3 shows baselines' accuracy scores for the pillar and stance classification tasks. We observe that the CLIP achieved the best performance on both tasks.

To further examine the predictions, we show example predictions of the best-performing model, CLIP, in Table 4. In example memes

(a) and (c), CLIP has correctly predicted the pillars and stances. More interestingly, we also noted that there are cases where the models made incorrect predictions in this task. For instance, in example meme (b), CLIP correctly predicted the "psychology" pillar but missed out on the "economic pillar". For example meme (d), CLIP correctly predicted the pillar but predicted an incorrect stance. The case studies and experimental results suggest room for improvement in the aspect-based stance classification task. We hope the release of TotalDefMeme can encourage more researchers to develop better aspect-based stance classification techniques.

6 CONCLUSION

In this paper, we have introduced TotalDefMeme, a novel multimodal multi-attribute meme dataset based on *Total Defence*. The dataset extends beyond the traditional hateful/offensive meme datasets by incorporating an aspect of public policy analysis from a social community perspective. We believe that social media memes capture a raw take on the bread and butter issues relating to governmental policies and initiatives, thereby providing the underlying public sentiments that may not be accurately captured from public surveys and questionnaires. This dataset is collected from the web without any preset limit on the number or domain of topics, and is annotated with the relevant defence pillars, associated topics and stance labels based on expert knowledge. Therefore, it provides the research community with a baseline for a comprehensive study on social media memes with respect to Total Defence or related governmental policies. We envisage the dataset to continue to push the boundaries for multimodal meme understanding and further the research for visual-language aspect-based stance classification and multimodal meme clustering.

REFERENCES

- [1] Rui Cao, Roy Ka-Wei Lee, Wen-Haw Chong, and Jing Jiang. 2023. Prompting for Multimodal Hateful Meme Classification. *arXiv preprint arXiv:2302.04156* (2023).
- [2] Anh Dang, Abidrahman Moh'd, Anatoliy Gruzd, Evangelos Milios, and Rosane Minghim. 2015. A visual framework for clustering memes in social media. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 713–720.

- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [4] Abhimanyu Dubey, Esteban Moro, Manuel Cebrian, and Iyad Rahwan. 2018. Memesequencer: Sparse matching for embedding image macros. In *Proceedings of the 2018 World Wide Web Conference*. 1225–1235.
- [5] Emilio Ferrara, Mohsen JafariAsbagh, Onur Varol, Vahed Qazvinian, Filippo Menczer, and Alessandro Flammini. 2013. Clustering memes in social media. In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*. IEEE, 548–555.
- [6] Ming Shan Hee, Roy Ka-Wei Lee, and Wen-Haw Chong. 2022. On Explaining Multimodal Hateful Meme Detection Models. In *Proceedings of the ACM Web Conference 2022*. 3651–3655.
- [7] Mohsen JafariAsbagh, Emilio Ferrara, Onur Varol, Filippo Menczer, and Alessandro Flammini. 2014. Clustering memes in social media streams. *Social Network Analysis and Mining* 4, 1 (2014), 1–13.
- [8] 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 (2020), 2611–2624.
- [9] Roy Ka-Wei Lee, Rui Cao, Ziqing Fan, Jing Jiang, and Wen-Haw Chong. 2021. Disentangling hate in online memes. In *Proceedings of the 29th ACM International Conference on Multimedia*. 5138–5147.
- [10] Jure Leskovec, Lars Backstrom, and Jon M. Kleinberg. 2009. Meme-tracking and the dynamics of the news cycle. In *Knowledge Discovery and Data Mining*.
- [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 preprint arXiv:1908.03557* (2019).
- [12] Lambert Mathias, Shaoliang Nie, Aida Mostafazadeh Davani, Douwe Kiela, Vinodkumar Prabhakaran, Bertie Vidgen, and Zeerak Waseem. 2021. Findings of the WOA5 Shared Task on Fine Grained Hateful Memes Detection. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*. Association for Computational Linguistics, Online, 201–206. <https://doi.org/10.18653/v1/2021.woah-1.21>
- [13] Shraman Pramanick, Md. Shad Akhtar, and Tanmoy Chakraborty. 2021. Exercise? I thought you said 'Extra Fries': Leveraging Sentence Demarcations and Multi-hop Attention for Meme Affect Analysis. *ArXiv abs/2103.12377* (2021).
- [14] Shraman Pramanick, Dimitar Dimitrov, Rituparna Mukherjee, Shivam Sharma, Md Akhtar, Preslav Nakov, Tanmoy Chakraborty, et al. 2021. Detecting harmful memes and their targets. *arXiv preprint arXiv:2110.00413* (2021).
- [15] Jingnong Qu, Liunian Harold Li, Jieyu Zhao, Sunipa Dev, and Kai-Wei Chang. 2022. DisinfoMeme: A Multimodal Dataset for Detecting Meme Intentionally Spreading Out Disinformation. *arXiv preprint arXiv:2205.12617* (2022).
- [16] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.
- [17] Chhavi Sharma, Deepesh Bhageria, William Scott, Srinivas PYKL, Amitava Das, Tanmoy Chakraborty, Viswanath Pulabaigari, and Björn Gambäck. 2020. SemEval-2020 Task 8: Memotion Analysis- the Visuo-Lingual Metaphor!. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. International Committee for Computational Linguistics, Barcelona (online), 759–773. <https://doi.org/10.18653/v1/2020.semeval-1.99>
- [18] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [19] Wee Yang Soh. 2020. Digital protest in Singapore: the pragmatics of political Internet memes. In *Sage Journals - Media, Culture & Society*, Vol. 42. 1115–1132. Issue 7-8.
- [20] 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*. European Language Resources Association (ELRA), Marseille, France, 32–41. <https://aclanthology.org/2020.trac-1.6>
- [21] William Theisen, Joel Brogan, Pamela Bilo Thomas, Daniel Moreira, Pascal Phoa, Tim Weninger, and Walter Scheirer. 2020. Automatic Discovery of Meme Genres with Diverse Appearances. *Association for the Advancement of Artificial Intelligence* (2020), 1–13.
- [22] Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. 2018. On the Origins of Memes by Means of Fringe Web Communities. In *Proceedings of the Internet Measurement Conference 2018 (Boston, MA, USA) (IMC '18)*. Association for Computing Machinery, New York, NY, USA, 188–202. <https://doi.org/10.1145/3278532.3278550>
- [23] Jiawen Zhu, Roy Ka-Wei Lee, and Wen Haw Chong. 2022. Multimodal zero-shot hateful meme detection. In *14th ACM Web Science Conference 2022*. 382–389.