Findings of Factify 2: Multimodal Fake News Detection

S Suryavardan*[1], Shreyash Mishra*[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]*University of California Los Angeles, USA*
[4]*Stanford, USA*
[5]*Amazon AI, USA*
[6]*Carnegie Mellon University, USA*
[7]*Microsoft, USA*
[8]*IIT Patna, India*
[9]*Georgia Tech, USA*

Abstract

With social media usage growing exponentially in the past few years, fake news has also become extremely prevalent. The detrimental impact of fake news emphasizes the need for research focused on automating the detection of false information and verifying its accuracy. In this work, we present the outcome of the Factify 2 shared task, which provides a multi-modal fact verification and satire news dataset, as part of the DeFactify 2 workshop at AAAI'23. The data calls for a comparison based approach to the task by pairing social media claims with supporting documents, with both text and image, divided into 5 classes based on multi-modal relations. In the second iteration of this task we had over 60 participants and 9 final test-set submissions. The best performances came from the use of DeBERTa for text and Swinv2 and CLIP for image. The highest F1 score averaged for all five classes was 81.82%.

Keywords

Fake News, Fact Verification, Multimodality, Dataset, Machine Learning, Entailment

1. Introduction

In recent years, automatic fact-checking has become an important problem in the AI community due to the increasing prevalence of fraudulent claims masquerading as declarations of reality. The rapid distribution of news across numerous media sources, particularly on social media

*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. 2023 Washington, DC, USA

suryavardan.s19@iiits.in (S. Suryavardan*); shreyash.m19@iiits.in (S. Mishra*); parthpatwa@g.ucla.edu (P. Patwa); amitava@mailbox.sc.edu (A. Das)

platforms, has led to the fast development of erroneous and fake content. The fake news becomes even more harmful during pandemic, elections etc [1, 2, 3, 4]. Uncovering misleading statements before they cause significant harm has become a challenging task. Studies indicate that a large percentage of the population believes that fake news creates uncertainty, while only a fraction feels confident in recognizing bogus news. However, the scarcity of available training data has hindered automated fact-checking efforts.

Significant progress has been made with the release of large datasets like [5, 6, 7], which include extensive claims and contextual metadata. Although these datasets have contributed to research advancements, they were purpose-made and may not capture the patterns present in real-world data effectively. Furthermore, most of the existing datasets focus only on text-based fake news. To address this limitation, Factify 1 [8] introduced a multimodal fact-checking dataset that consists of original samples with no post-processing or manual data creation involved. The dataset includes images, textual claims, and reference textual documents/images, facilitating the exploration of visual cues to enhance fake content detection.

Factify 2 [9], the latest iteration of the Factify dataset, introduced several enhancements and new challenges. Factify 2 expands the dataset with an additional 50,000 instances, encompassing satirical articles that present fake news in a different manner. By incorporating satirical content, it aims to capture the nuanced complexities associated with detecting misinformation in diverse formats. The Factify datasets addresses the limitations of previous unimodal fact-checking research efforts by providing a benchmark for researchers to build and evaluate multimodal fact verification systems. Each data sample in the dataset is labeled with one of five choices: support, no-evidence, refute (both in text and multimodal components), and satirical. The data is collected from popular news channels' Twitter handles in the United States and India, ensuring its relevance to real-world scenarios. Factify 2 builds upon the success of its predecessor by expanding the dataset, introducing satirical articles, and creating new challenges for multimodal fact verification.

This paper presents the findings of the shared task Factify 2, which was organized as part of the workshop at AAAI 2023. The shared task brought together researchers and practitioners to evaluate and compare their approaches in multimodal fact verification using the Factify 2 dataset. Section 2 describes the related work in this domain. We detail the task and the requirements in 3, followed by the participating teams in section 4. The results of all participating teams' models are presented in Section 5. Finally, we summarize the task, discuss further research opportunities, and provide open-ended pointers in Section 7.

2. Related Work

Multi-modal learning: Researchers have looked into a number of models that incorporate both textual and visual information in the field of multi-modal learning. This includes approaches that range from concatenation of individual embeddings to attention fusion layers. ViLBERT [10] is one such approach that processes both visual and textual inputs separately. They interact through co-attentional transformer layers while extending the BERT architecture. LXMERT [11] builds a large-scale Transformer model with an object relationship encoder, a language encoder, and a cross-modality encoder, for which the model is trained via pre-training tasks like masked language modeling, masked object prediction, etc. to learn intra and cross modality relationships. Learning Joint embedddings using object-word pairs is proposed by [12, 13]. Other notable work includes OFA [14], BLIP [15], ALIGN [16], CLIP[17], etc.

Unimodal fake news detection: Many workshops and shared tasks have been conducted on unimodal/text based fake news detection and fact verification [18, 19]. Datasets like LIAR [5], Covid-19 fake news [20], CREDBANK [21] use claims from sources such as social media and PolitiFact and are labeled into fake/real or other fine-grained categories. FEVER [18] provides claims with supporting wikipedia articles that are classified into ”Supported”, ”Refuted” or ”NotEnoughInfo”. Fever and Feverous[22] use artificial claims that were generated or modified for the task, which may not be useful in real world scenarios. Recently, bert-based methods have been very popular in to tackle text based fake news detection [19, 23, 24]. While unimodal fake news detection might be a good starting point, there is information loss by not utilizing multiple modalities, which may be critical in fake news detection.

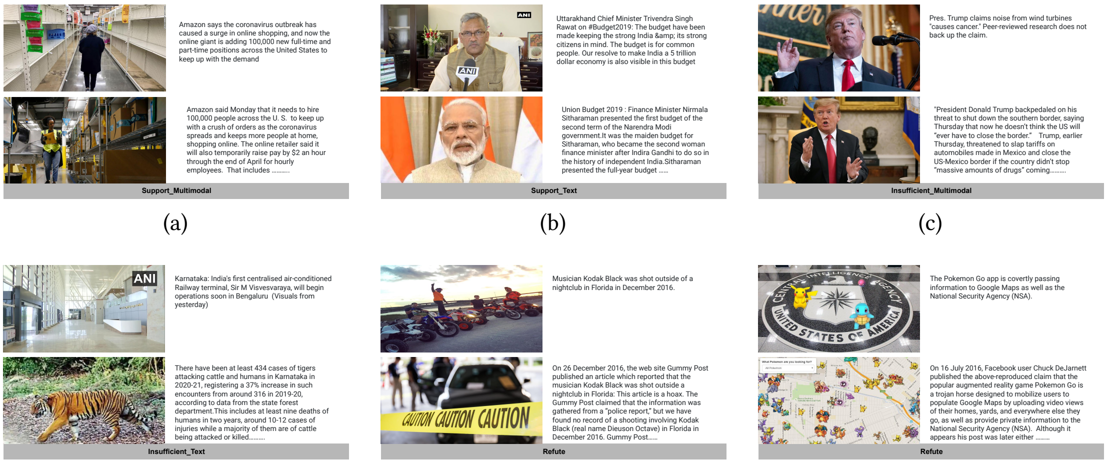
Multimodal fake news detection: Although under-explored, some of the recent works aim for multimodal fake news detection. The Fakeddit dataset [25] has 1 million samples containing text-image pairs classified into fine-grained labels as well as high-level labels. FakeNewsNet [26], collects data from fact checking websites focused on mitigation and spreading for fake news, and provide social context and dynamic information along with the news. Papadopoulou et al. [27] provide a dataset of 380 verified and debunked videos. Methods to tackle this problem have been proposed in [28, 29, 30, 31]. For a detailed survey on multimodal fake news detection, please refer [32].

Factify 1: The previous iteration, Factify 1 shared task [33] at AAAI 2022 was conducted on a multi-modal dataset of having 50k instances. Each data point includes image-text pair for a claim and supporting document, and each claim-document pair is classified into Support, insufficient, neutral. Factify 2 releases additional 50k instances which include data points taken from satirical news.

3. Task Details

The Factify 2 task is designed similar to the previous iteration of the task [8]. The task aims to detect fake news with an automated model that is able to classify a given claim based on whether the document entails it or refutes it. Thus, the dataset has 50,000 data points such that each sample has a claim-document pair. Here, the claim is defined as a social media post and the document is an article surrounding the claim. Due to the nature of social media content, the dataset encourages the use of textual and visual features by providing a text and an image for every claim and document. The entailment between the four data sources, namely claim image, claim text, document image and document text, are used to define the categories that the data is to be classified into. This is also shown in Figure 2.

The task description and access to the dataset is available in the Factify 2 task page at https://codalab.lisn.upsaclay.fr/competitions/8275.



Amazon says the coronavirus outbreak has caused a surge in online shopping, and now the online giant is adding 100,000 new full-time and part-time positions across the United States to keep up with the demand

Amazon said Monday that it needs to hire 100,000 people across the U. S. to keep up with a crush of orders as the coronavirus spreads and keeps more people at home, shopping online. The online retailer said it will also temporarily raise pay by $2 an hour through the end of April for hourly employees. That includes ....

(a)

Uttarakhand Chief Minister Trivendra Singh Rawat on #Budget2019: The budget have been made keeping the strong India &amp; its strong citizens in mind. The budget is for common people. Our resolve to make India a 5 trillion dollar economy is also visible in this budget

Union Budget 2019 : Finance Minister Nirmala Sitharaman presented the first budget of the second term of the Narendra Modi government.It was the maiden budget for Sitharaman, who became the second woman finance minister after Indira Gandhi to do so in the history of independent India.Sitharaman presented the full-year budget ......

(b)

Pres. Trump claims noise from wind turbines "causes cancer." Peer-reviewed research does not back up the claim.

"President Donald Trump backpedaled on his threat to shut down the southern border, saying Thursday that now he doesn't think the US will "ever have to close the border." Trump, earlier Thursday, threatened to slap tariffs on automobiles made in Mexico and close the US-Mexico border if the country didn't stop "massive amounts of drugs" coming..........

(c)

Karnataka: India's first centralised air-conditioned Railway terminal, Sir M Visvesvaraya, will begin operations soon in Bengaluru (Visuals from yesterday)

There have been at least 434 cases of tigers attacking cattle and humans in Karnataka in 2020-21, registering a 37% increase in such encounters from around 316 in 2019-20, according to data from the state forest department.This includes at least nine deaths of humans in two years, around 10-12 cases of injuries while a majority of them are of cattle being attacked or killed.........

(d)

Musician Kodak Black was shot outside of a nightclub in Florida in December 2016.

On 26 December 2016, the web site Gummy Post published an article which reported that the musician Kodak Black was shot outside a nightclub in Florida: This article is a hoax. The Gummy Post claimed that the information was gathered from a "police report," but we have found no record of a shooting involving Kodak Black (real name Dieuson Octave) in Florida in December 2016. Gummy Post......

(e)

The Pokemon Go app is covertly passing information to Google Maps as well as the National Security Agency (NSA).

On 16 July 2016, Facebook user Chuck DeJarnett published the above-reproduced claim that the popular augmented reality game Pokemon Go is a trojan horse designed to mobilize users to populate Google Maps by uploading video views of their homes, yards, and everywhere else they go, as well as provide private information to the National Security Agency (NSA). Although it appears his post was later either ....

CENTRAL INTELLIGENCE AGENCY
UNITED STATES OF AMERICA
What Pokemon are you looking for?
All Pokemon
Refute

(f)

Figure 1: Examples for all the 5 categories.

Support_Multimodal	Text is supported Similar News	Image is supported
Support_Text	Text is supported Similar News	Image is neither supported nor refuted
Insufficient_Multimodal	Text is neither supported nor refuted May have common words	Image is supported
Insufficient_Text	Text is neither supported nor refuted May have common words	Image is neither supported nor refuted
Refute	Fake Claim	Fake Image

Figure 2: This figure shows the five categories the dataset has been divided into along with the relationship between the multi-modal claim and document in each class.

3.1. Data

The dataset [9] consists of claim-document pairs and is curated by combining data from Twitter, Fact Checking websites and Satirical news websites 1. The claims were extracted from tweets from Hindustan Times, ANI, ABC and CNN, with their corresponding document text extracted from the news articles linked to the tweets. Based on metrics like textual and image similarity, the collected samples were classified into the Support and Neutral categories. Samples for the refute category were collected from fact checking websites - the fake news was selected as the claim and the article contents were chosen as the corresponding document. Similarly, Satirical

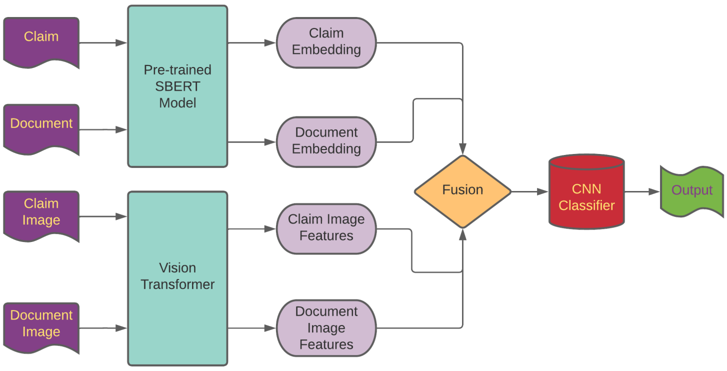


Figure 3: This is the baseline model architecture. It extracts multi-modal features using pre-trained models, followed by their fusion and subsequent classification.

websites were scraped for text and image, and as the article supports the satirical i.e. false claim, they were added to the Support category. Satirical headlines were also searched on Google to find news articles that Refuted or were Insufficient for the claim. The train-validation-test split of the 50,000 samples is 70:15:15, with each category having the same number of samples. The train and validation split was provided to the participants and the test set was hidden and used for evaluation. A more in-depth description of the collection process is presented in the data paper for the task [9].

3.2. Baseline

The baseline for the Factify 2 dataset is presented in [9]. It uses a pre-trained textual and visual feature extraction models before concatenating the features and passing it to a classifier, as shown in Figure 3. The data paper compares few pre-trained models for feature extraction and we finds that the SBERT-mpnet model for text and ViT for image gave the best performance with a F1 score of 64.99%. Please refer to [9] for more details of the baseline.

3.3. Evaluation

The Factify 2 dataset has 5 classes, that the data points are to be categorised into: “Support_Text“, “Support_Multimodal“, “Insufficient_Text“, “Insufficient_Multimodal“ and “Refute“. To compute the performance of the classification models, we use F1 score between the ground truth labels and the predicted labels.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}, Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

We calculate the weighted average F1 score for individual classes and also for all samples together. These scores are computed for all the participating systems and the baseline, as shown in Table 1. The participants were allowed to make 3 submission and the best score out of them was used for the leaderboard. The final F1 score across all classes is used to rank the participants.

4. Participating systems

We had over 60 registrations, of which 9 teams provided their final predictions on the test set and 7 teams made paper submissions to the workshop. We provide an overview of the approaches used by the participating teams in this section.

Triple-Check [34] propose a model with pre-trained DeBERTa[35] for text and Swinv2[36] for image embeddings, that are combined using a co-attention fusion block. Their novelty is in the use of an adapter to train only a few parameters in their model. Features such as text length, OCR etc. are also used for the final classification.

INO [37] use a structure coherence-based approach with components such as textual feature similarity, textual semantic similarity, text length and image similarity. CLIP [17], S-BERT [38] and the ROUGE [39] are used for text features and ResNet for image features. These components are used for the final classification through a Random forest classifier.

Logically [40] call their architecture a cross-modal veracity prediction model. They use both uni-modal embeddings from Word2Vec and multi-modal embeddings from CLIP for their classification, after passing them through a multi-head attention layer. The document being fed to this model is from a evidence-retrieval stage, where they rank the paragraphs in document and use the top-K passages.

Zhang [41] use uni-modal embeddings from pre-trained DeBERTa and ViT, passed to signed attention layers and a feed-forward network, as well as multi-modal embeddings from CLIP for their classification. This defined as three modules, namely a text-semantic feature module, a image-semantic feature module and a text-image correlation module, together referred to as 'UFCC'.

gzw [42] experimented with a inter-modality and intra-modality fusion of textual and visual embeddings using the co-attention mechanism for their classification model. They refer to this architecture as Multimodal Attention and Fusion Network (MAFN). They use different pre-trained models within MAFN and make predictions with an ensemble.

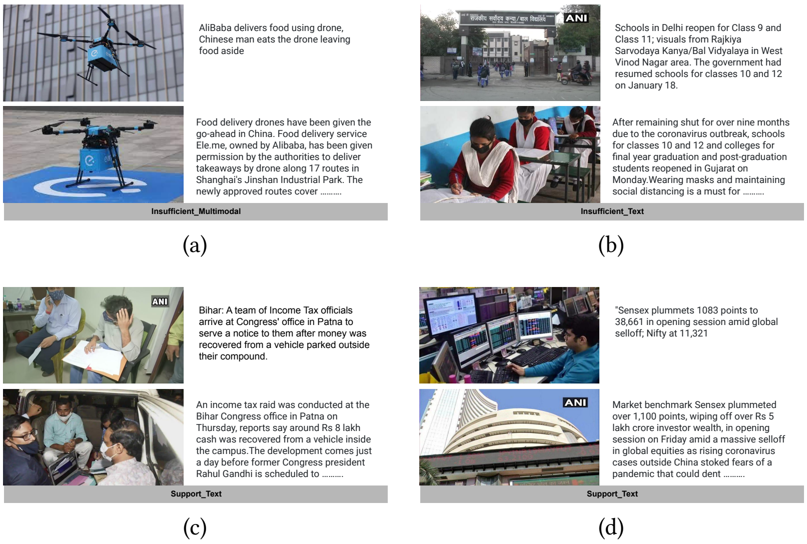
coco [43] use an embeddings layer to combine individual embeddings from pre-trained DeBERTa for text and DeiT[44] for images. This embedding is then passed to bi-directional hybrid attention mechanism to fuse the claim and document data. The final classification uses an ensemble of two model pipelines.

Noir [45] obtain image features from pre-trained CLIP-ViT and text features from multilingual CLIP-ViT. Claim and document embeddings are concatenated before being fed to a multi-headed attention layer. Then they train a MLP for classification with the output features for image and text concatenated.

5. Results

Rank	Team	Support_Text	Support_Multimodal	Insufficient_Text	Insufficient_Multimodal	Refute	Final
1	**Triple-Check** [34]	**82.77%**	**91.38%**	85.19%	**89.22%**	**100.0%**	**81.82%**
2	INO [37]	81.24%	90.03%	**88.81%**	85.23%	99.93%	80.80%
3	Logically [40]	80.38%	90.51%	84.39%	85.63%	98.51%	78.97%
4	zhang [41]	76.64%	87.85%	81.61%	87.93%	99.93%	77.42%
5	gzw [42]	78.49%	86.32%	81.42%	83.27%	**100.0%**	76.05%
6	coco [43]	77.25%	86.49%	81.52%	83.00%	**100.0%**	75.70%
7	Noir [45]	77.10%	87.26%	78.49%	81.56%	99.70%	74.52%
8	Yet	70.75%	82.63%	78.59%	71.90%	**100.0%**	69.09%
-	BASELINE	50.00%	82.72%	80.24%	75.93%	98.82%	64.99%
9	TeamX	58.22%	70.91%	53.66%	55.56%	69.79%	45.62%

Table 1
Leaderboard for the Factify task. The teams are ranked by the overall weighted average F1 score (`Final` column). We also report the category-wise weighted avg F1 score for each team.


AliBaba delivers food using drone, Chinese man eats the drone leaving food aside

Food delivery drones have been given the go-ahead in China. Food delivery service Ele.me, owned by Alibaba, has been given permission by the authorities to deliver takeaways by drone along 17 routes in Shanghai's Jinshan Industrial Park. The newly approved routes cover ....

(a)

Schools in Delhi reopen for Class 9 and Class 11; visuals from Rajkiya Sarvodaya Kanya/Bal Vidyalaya in West Vinod Nagar area. The government had resumed schools for classes 10 and 12 on January 18.

After remaining shut for over nine months due to the coronavirus outbreak, schools for classes 10 and 12 and colleges for final year graduation and post-graduation students reopened in Gujarat on Monday.Wearing masks and maintaining social distancing is a must for ....

(b)

Bihar: A team of Income Tax officials arrive at Congress' office in Patna to serve a notice to them after money was recovered from a vehicle parked outside their compound.

An income tax raid was conducted at the Bihar Congress office in Patna on Thursday, reports say around Rs 8 lakh cash was recovered from a vehicle inside the campus.The development comes just a day before former Congress president Rahul Gandhi is scheduled to ....

(c)

"Sensex plummets 1083 points to 38,661 in opening session amid global selloff; Nifty at 11,321

Market benchmark Sensex plummeted over 1,100 points, wiping off over Rs 5 lakh crore investor wealth, in opening session on Friday amid a massive selloff in global equities as rising coronavirus cases outside China stoked fears of a pandemic that could dent ....

(d)

Figure 4: Examples from the dataset that were incorrectly predicted by all participants of the task. The Support_Text category has the highest number of such samples, followed by Insufficient_Multimodal, Support_Multimodal, Insufficient_Text and finally Refute in descending order.

For Factify 2, the participants were allowed to make a maximum of three submissions on the test set. The Final F1 score was used to decide which of the three prediction files is the best performing for each participant. Based on that criteria, we present the results of the 9 teams that made test-phase submissions in Table 1. The table shows the weighted average F1 score for the 5 individual classes as well as for the overall test set. All teams except one, improved on the baseline by a minimum of 6.3%. The best performing team is Triple-Check [34] with a Final Score of 81.82%, which is about 26% higher than the baseline. No single team performed better than other team in all categories, which shows that the problem is challenging and needs further research attention.

Despite adding the satire news articles, the results for the Refute category are very high, similar to the Factify 1. This may be due to the fact that the document collected from of fact checking websites are different to the writing in formal news articles. While only having a marginal difference, Support_Text has the lowest scores compared to other categories. This is further emphasised by the observation that, of samples that were predicted incorrectly by all participants, Support_Text has the highest occurrence, while Refute followed by Insufficient_Text occur the least. There are some confusing data points where all the systems failed to predict the correct class. Some such examples are displayed in Figure 4. From the first example, we can see that the document is unrelated to claim, but both of them have a similar drone image, hence leading to confusion. Similarly, in the second example, both the document and claim are about schools re-opening but in different locations, which the models failed to capture.

6. Conclusion and Future Work

In this paper, we outline and discuss the Factify 2 shared task on multi-modal fact verification. The participants utilized various techniques for text embeddings, including DeBERTa, CLIP, S-BERT, ROUGE and Word2Vec. The image embeddings were extracted through Swinv2, ResNet, CLIP, ViT and DeiT. Similar to Factify1, ensemble techniques were popular, and some teams opted to use multiple embeddings to capture features. The shared task described in this work seeks to identify fake news, but we are far from our goal. No single system could excel in all categories and there are few examples where all the systems failed, which highlights the challenges of the task. One possible direction for further research includes using synthetic fake news data that matches the general data distribution, thus adding complexity to the refute category.

References

[1] N. Karimi, J. Gambrell, hundreds die of poisoning in iran as fake news suggests methanol cure for virus, 2020. URL: https://www.timesofisrael.com/hundreds-die-of-poisoning-in-iran-as-fake-news-suggests-methanol-cure-for-virus/.

[2] J. Bae, D. Gandhi, J. Kothari, S. Shankar, J. Bae, P. Patwa, R. Sukumaran, A. Chharia, S. Adhikesaven, S. Rathod, I. Nandutu, S. TV, V. Yu, K. Misra, S. Murali, A. Saxena, K. Jakimowicz, V. Sharma, R. Iyer, A. Mehra, A. Radunsky, P. Katiyar, A. James, J. Dalal, S. Anand, S. Advani, J. Dhaliwal, R. Raskar, Challenges of equitable vaccine distribution in the covid-19 pandemic, 2022. arXiv:2012.12263.

[3] M. Morales, R. Barbar, D. Gandhi, S. Landage, J. Bae, A. Vats, J. Kothari, S. Shankar, R. Sukumaran, H. Mathur, K. Misra, A. Saxena, P. Patwa, S. T. V., M. Arseni, S. Advani, K. Jakimowicz, S. Anand, P. Katiyar, A. Mehra, R. Iyer, S. Murali, A. Mahindra, M. Dmitrienko, S. Srivastava, A. Gangavarapu, S. Penrod, V. Sharma, A. Singh, R. Raskar, Covid-19 tests gone rogue: Privacy, efficacy, mismanagement and misunderstandings, 2021. arXiv:2101.01693.

[4] S. Muhammed T, S. K. Mathew, The disaster of misinformation: a review of research in social media, International Journal of Data Science and Analytics 13 (2022) 271–285.

[5] W. Y. Wang, " liar, liar pants on fire": A new benchmark dataset for fake news detection, arXiv preprint arXiv:1705.00648 (2017).

[6] T. Mitra, E. Gilbert, Credbank: A large-scale social media corpus with associated credibility annotations, in: ICWSM, 2015.

[7] I. Augenstein, C. Lioma, D. Wang, L. Chaves Lima, C. Hansen, C. Hansen, J. Grue Simonsen, Multifc: A real-world multi-domain dataset for evidence-based fact checking of claims, in: EMNLP, Association for Computational Linguistics, 2019.

[8] S. Mishra, S. Suryavardan, A. Bhaskar, P. Chopra, A. Reganti, P. Patwa, A. Das, T. Chakraborty, A. Sheth, A. Ekbal, et al., Factify: A multi-modal fact verification dataset, in: De-Factify workshop at AAAI, 2022.

[9] S. Suryavardan, S. Mishra, P. Patwa, M. Chakraborty, A. Rani, A. Reganti, A. Chadha, A. Das, A. Sheth, M. Chinnakotla, A. Ekbal, S. Kumar, Factify 2: A multimodal fake news and satire news dataset, in: De-Factify 2 workshop at AAAI, 2023.

[10] J. Lu, D. Batra, D. Parikh, S. Lee, Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks, Advances in neural information processing systems 32 (2019).

[11] H. Tan, M. Bansal, Lxmert: Learning cross-modality encoder representations from transformers, 2019. arXiv:1908.07490.

[12] 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) 12959–12960. URL: https://ojs.aaai.org/index.php/AAAI/article/view/21616. doi:10.1609/aaai.v36i11.21616.

[13] 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.

[14] P. Wang, A. Yang, R. Men, J. Lin, S. Bai, Z. Li, J. Ma, C. Zhou, J. Zhou, H. Yang, Ofa:

Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework, 2022. arXiv:2202.03052.

[15] J. Li, D. Li, C. Xiong, S. Hoi, Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation, 2022. arXiv:2201.12086.

[16] C. Jia, Y. Yang, Y. Xia, Y.-T. Chen, Z. Parekh, H. Pham, Q. V. Le, Y. Sung, Z. Li, T. Duerig, Scaling up visual and vision-language representation learning with noisy text supervision, 2021. arXiv:2102.05918.

[17] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever, Learning transferable visual models from natural language supervision, 2021. arXiv:2103.00020.

[18] J. Thorne, A. Vlachos, C. Christodoulopoulos, A. Mittal, Fever: a large-scale dataset for fact extraction and verification, arXiv preprint arXiv:1803.05355 (2018).

[19] P. Patwa, M. Bhardwaj, V. Guptha, G. Kumari, S. Sharma, S. PYKL, A. Das, A. Ekbal, S. Akhtar, T. Chakraborty, Overview of constraint 2021 shared tasks: Detecting english covid-19 fake news and hindi hostile posts, in: Proceedings of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT), Springer, 2021.

[20] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, T. Chakraborty, Fighting an infodemic: Covid-19 fake news dataset, in: Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT) 2021, Springer, 2021, p. 21–29. URL: http://dx.doi.org/10.1007/978-3-030-73696-5_3. doi:10.1007/978-3-030-73696-5_3.

[21] T. Mitra, E. Gilbert, Credbank: A large-scale social media corpus with associated credibility annotations, in: Proceedings of the international AAAI conference on web and social media, volume 9, 2015, pp. 258–267.

[22] R. Aly, Z. Guo, M. S. Schlichtkrull, J. Thorne, A. Vlachos, C. Christodoulopoulos, O. Cocarascu, A. Mittal, The fact extraction and verification over unstructured and structured information (feverous) shared task, in: Proceedings of the Fourth Workshop on Fact Extraction and VERification (FEVER), 2021, pp. 1–13.

[23] B. Chen, B. Chen, D. Gao, Q. Chen, C. Huo, X. Meng, W. Ren, Y. Zhou, Transformer-based language model fine-tuning methods for covid-19 fake news detection, 2023. arXiv:2101.05509.

[24] R. K. Kaliyar, A. Goswami, P. Narang, Fakebert: Fake news detection in social media with a bert-based deep learning approach, Multimedia Tools and Applications 80 (2021) 11765–11788. URL: https://doi.org/10.1007/s11042-020-10183-2. doi:10.1007/s11042-020-10183-2.

[25] K. Nakamura, S. Levy, W. Y. Wang, r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection, arXiv preprint arXiv:1911.03854 (2019).

[26] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media, 2019. arXiv:1809.01286.

[27] O. Papadopoulou, M. Zampoglou, S. Papadopoulos, I. Kompatsiaris, A corpus of debunked and verified user-generated videos, Online information review 43 (2019) 72–88.

[28] Y. Wu, P. Zhan, Y. Zhang, L. Wang, Z. Xu, Multimodal fusion with co-attention networks

for fake news detection, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, Online, 2021, pp. 2560–2569. URL: https://aclanthology.org/2021.findings-acl.226. doi:10.18653/v1/2021.findings-acl.226.

[29] J. Jing, H. Wu, J. Sun, X. Fang, H. Zhang, Multimodal fake news detection via progressive fusion networks, Information processing & management 60 (2023) 103120.

[30] J. Hua, X. Cui, X. Li, K. Tang, P. Zhu, Multimodal fake news detection through data augmentation-based contrastive learning, Applied Soft Computing 136 (2023) 110125.

[31] A. Yadav, S. Gaba, H. Khan, I. Budhiraja, A. Singh, K. K. Singh, Etma: Efficient transformer-based multilevel attention framework for multimodal fake news detection, IEEE Transactions on Computational Social Systems (2023).

[32] F. Alam, S. Cresci, T. Chakraborty, F. Silvestri, D. Dimitrov, G. D. S. Martino, S. Shaar, H. Firooz, P. Nakov, A survey on multimodal disinformation detection, in: Proceedings of the 29th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 6625–6643. URL: https://aclanthology.org/2022.coling-1.576.

[33] P. Patwa, S. Mishra, S. Suryavardan, A. Bhaskar, P. Chopra, A. Reganti, A. Das, T. Chakraborty, A. Sheth, A. Ekbal, et al., Benchmarking multi-modal entailment for fact verification, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[34] W.-W. Du, H.-W. Wu, W.-Y. Wang, W.-C. Peng, Team Triple-check at factify 2: Parameter-efficient large foundation models with feature representations for multi-modal fact verification, in: De-Factify 2 workshop at AAAI, 2023.

[35] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, 2021. arXiv:2006.03654.

[36] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, B. Guo, Swin transformer v2: Scaling up capacity and resolution, 2022. arXiv:2111.09883.

[37] Y. Zhang, Z. Tao, X. Wang, T. Wang, INO at Factify 2: Structure coherence based multi-modal fact verification, in: De-Factify 2 workshop at AAAI, 2023.

[38] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, 2019. arXiv:1908.10084.

[39] C.-Y. Lin, ROUGE: A package for automatic evaluation of summaries, in: Text Summarization Branches Out, Association for Computational Linguistics, Barcelona, Spain, 2004, pp. 74–81. URL: https://aclanthology.org/W04-1013.

[40] P. Jordi Verschuuren, J. Gao, A. Van Eeden, S. Oikonomou, A. Bandhakavi, Logically at Factify 2: A multi-modal fact checking system based on evidence retrieval techniques and transformer encoder architecture, in: De-Factify 2 workshop at AAAI, 2023.

[41] F. Zhang, Q. Zhang, J. Liu, E. Sun, Team Zhang at Factify 2: Unimodal feature-enhanced and cross-modal correlation learning for multi-modal fact verification, in: De-Factify 2 workshop at AAAI, 2023.

[42] Z. Gao, Z. Wang, T. Chen, Team gzw at Factify 2023: Multimodal attention and fusion networks for multi-modal fact verification, in: De-Factify 2 workshop at AAAI, 2023.

[43] K. Guo, S. Luo, R. Ma, Y. Wang, Y. Zhang, Team coco at Factify 2: Optimizing modality representation and multimodal fusion for multimodal fact verification, in: De-Factify 2

workshop at AAAI, 2023.

[44] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jégou, Training data-efficient image transformers & distillation through attention, 2021. arXiv:2012.12877.

[45] Z. Zhang, H. Yang, C. Huang, Y. Zhang, Team Noir at Factify 2: Multimodal fake news detection with pre-trained clip and fusion network, in: De-Factify 2 workshop at AAAI, 2023.