TeamX at CheckThat! 2023: Multilingual and Multimodal Approach for Check-Worthiness Detection

Rabindra Nath Nandi[1,*], Firoj Alam[2,*]

[1]Hishab Singapore Pte. Ltd, Singapore
[2]Qatar Computing Research Institute, HBKU, Qatar

Abstract
Check-worthiness detection is a crucial aspect of the fact-checking pipeline. It aids fact-checkers and journalists by highlighting the claims that necessitate verification. This is especially pertinent in today's era of social media and varied news channels, where numerous actors voice claims on a range of current affairs topics, including political matters, global warming, and the COVID-19 vaccine. Furthermore, during political events, politicians debate their political agendas and make numerous claims on various subjects. These claims, which often lack factual basis, come in all forms. While some of these claims are important, others are not. Given the time-intensive nature of manual fact-checking, the identification of claims worthy of fact-checking becomes critical. Over the years, there have been research efforts aimed at the automatic detection of such claims. To further this research, in past years, `CheckThat! Lab` has offered check-worthiness detection tasks on political debates and textual modality social media content. For the first time, `CheckThat! Lab` offered the task for multimodal content. This study reports our participation in subtask-1A, which consists of Arabic and English. For our experiments, we utilized transformer-based models for both unimodal and multimodal models. The performances of the submitted systems, evaluated using the F1-score on the positive class, were 0.671 and 0.300, respectively. Our systems did not rank on the leaderboard as we made late submissions. However, with additional experiments, we achieved 0.684 and 0.362 for English and Arabic, respectively.

Keywords
Check-worthiness, Check-worthy claim detection, Fact-checking, Disinformation, Misinformation, Social Media Text, Transformer Models

1. Introduction

Claim detection plays a significant role in both argument mining [1] and automated fact-checking pipelines [2]. The objective is to assist fact-checkers and journalists in their fact-checking processes. While earlier work on fact-checking primarily focused on political debates, the increasing prevalence of misleading information shared via social media and other news channels has become problematic. As a result, the analysis of social media content has garnered significant attention [3, 4].

Manual fact-checking has traditionally been the norm for verifying claims. As a result,

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece
*Corresponding author.
rabindro.rath@gmail.com (R. N. Nandi); fialam@hbku.edu.qa (F. Alam)

many fact-checking organizations, such as *FactCheck.org*,[1], *Snopes*,[2], *PolitiFact*,[3], and *FullFact*[4], have emerged over time. Although manual fact-checking is trustworthy, it does not scale well for several reasons: the task of identifying important claims to fact-check and verifying the truthfulness of a claim with evidence. Fact-checkers must prioritize the evaluation of claims that are potentially harmful and can pose risks to health, democratic processes, or exacerbate emergency situations. Considerable research effort has been devoted to identifying such significant claims, thereby streamlining the fact-checkers' manual effort [5, 6].

Over the past few years, the `CheckThat! Lab` initiative has been promoting the development of systems for detecting check-worthiness in political debates, tweets, and transcripts [7, 8, 9, 10]. This year, the `CheckThat! Lab` has introduced five tasks covering seven languages (Arabic, Dutch, English, German, Italian, Spanish, and Turkish). It includes content from various genres and modalities [11].

We participated in the check-worthiness task, which focused on multimodal and multigenre content in three different languages: Arabic, English, and Spanish [11, 12, 13]. This task comprised two subtasks – Multimodal (1A) and Multigenre (1B). We took part in subtask 1A.

In our experiments, we utilized various unimodal and multimodal transformer-based models. Our submitted system outperformed three other systems, including the baseline, for English, and it also surpassed the Arabic baseline. However, due to our late submission, our systems did not secure a place on the leaderboard. After the gold labels for the test set were released, additional experiments allowed us to match the performance of the best system for Arabic and achieve improved results for English.

The rest of this paper is structured as follows. In Section 2, we discuss the related works that are relevant to this study. The methodology is detailed in Section 3. The results of the experiments, along with in-depth discussions, are provided in Section 4. Finally, we conclude our study in Section 5.

2. Related Work

Within the scope of identifying disinformation, misinformation, and "fake news" in general, research interests have focused on more specific problems. These range from automatic identification and verification of claims [14, 15, 16], and identifying check-worthy claims [17, 18, 19, 20, 21], to detecting whether a claim has been previously fact-checked [22, 2, 23]. Other areas of focus include retrieving evidence to accept or reject a claim [24], checking whether the evidence supports or denies the claim [25, 26], and inferring the veracity of the claim [27, 28, 29].

Among these tasks, check-worthiness estimation has received wider attention since the pioneering work proposed by [30]. The aim is to detect whether a sentence in a political debate is *non-factual*, *unimportant factual*, or *check-worthy factual*. This proposed system was later extended with more data and modified to cover Arabic content [17]. Most of the earlier work on check-worthiness estimation primarily focused on political debates [31, 18], but lately, attention

[1]http://www.factcheck.org
[2]http://www.snopes.com/fact-check/
[3]http://www.politifact.com
[4]http://fullfact.org

Table 1
Data splits and distributions of the dataset for Subtask 1A: Check-Worthiness detection from multimodal content.

Class labels	Train	Dev	Dev-Test	Test	Total
Arabic (Multimodal)					
No	1,421	207	402	792	2,822
Yes	776	113	220	203	1,312
Total	2,197	320	622	995	4,134
English (Multimodal)					
No	1,536	184	374	459	2,553
Yes	820	87	174	277	1,358
Total	2,356	271	548	736	3,911

has been directed towards social media [3, 32, 33].

Significant research attention has emerged due to the `CheckThat! Lab` initiatives started back in CLEF 2018 [34, 35, 36, 37]. The focus, once again, was on political debates and speeches from a single fact-checking organization. In the 2018 edition of the task, a total of seven teams submitted runs for Task1 (which corresponds to subtask-1B in 2021), with systems based on word embeddings and RNNs. In the 2021 edition [38], check-worthiness estimation was offered for both political debates/speeches and tweets, while in the 2022 edition, it was offered only for tweets [20].

3. Experiments

3.1. Tasks and Datasets

The aim of this task is to determine whether a claim is worth fact-checking. It has been offered in two subtasks: *(i)* subtask 1A (multimodal), where each instance comprises the text and the image associated with a tweet, and *(ii)* subtask 1B, where each instance consists of only text, derived from a tweet, the transcription of a debate, or the transcription of a speech. Subtask 1A is offered in Arabic and English, while subtask 1B is available in Arabic, English, and Spanish. As previously mentioned, our study primarily focuses on subtask 1A (multimodal).

We used the dataset provided by `CheckThat! Lab`. The distribution of the dataset for subtask 1A is shown in Table 1. The dataset for the development phase consists of train, dev, and dev-test sets, and an additional test set without gold labels was provided during the evaluation phase. During the development phase, we utilized the train and dev sets for training and fine-tuned the models and used the dev-test set for the evaluation of the systems. The dev-test set is considered as a held-out set in this phase. During the evaluation phase, we classified the test set and submitted it for the evaluation.

Exp	Dataset	Acc	P	R	F1
Baseline and submitted systems					
Baseline	AR				0.299
Baseline	EN				0.474
Our submission (ViT + araBERT)	AR				0.300
Our submission (ViT + mBERT (Tweets) + mBERT (OCR))	EN				0.671
Text modality					
araBERT	AR	0.673	0.319	0.532	**0.399**
mBERT	EN	0.789	0.765	0.635	**0.694**
Image modality					
ResNet18	AR	0.601	0.422	0.166	0.238
ResNet101	AR	0.615	0.464	0.141	0.216
Vgg16	AR	0.594	0.423	0.217	0.286
Efficientnet (b1)	AR	0.599	0.446	0.267	0.334
Efficientnet (b7)	AR	0.601	0.442	0.235	0.307
ResNet18	EN	0.636	0.529	0.292	0.377
ResNet101	EN	0.641	0.536	0.350	0.424
Vgg16	EN	0.636	0.529	0.292	0.377
Efficientnet (b1)	EN	0.611	0.473	0.289	0.359
Efficientnet (b7)	EN	0.599	0.462	0.394	0.425
Multimodality					
ViT + araBERT	AR	0.661	0.294	0.473	0.362
ViT + mBERT (Tweets)	EN	0.776	0.750	0.606	0.671
ViT + mBERT (Tweets) + Add. data	EN	0.783	0.755	0.625	0.684

Table 2
Evaluation results on the test set. Best results are highlighted in bold.

3.2. Settings

For the classification experiments, we trained different unimodal and multimodal models, which involved using *(i)* only text, *(ii)* only images, and *(iii)* both text and images together.

3.2.1. Text Modality

For the text modality experiment, we fine-tuned transformer models. For the English dataset, we trained the model using the multilingual BERT (mBERT) model [39], and for Arabic, we used the araBERT model [40]. We fine-tuned the models and selected the one that performed the best on the development set. We used a batch size of 8, a learning rate of 1e-6, a maximum sequence length of 512, a maximum of 50 epochs, early stopping, and employed categorical cross-entropy as the loss function.

3.2.2. Image Modality

For the image modality experiments, we employed the transfer learning approach by fine-tuning pre-trained deep CNN models such as VGG16, which has demonstrated success in visual recognition tasks [41]. We utilized the weights of the model pre-trained on ImageNet to initialize our model. We adapted the last layer (i.e., softmax layer) of the network for the binary classification task. Our models were trained using three popular neural network architectures: VGG16 [42], ResNet101 [43], and EfficientNet [44], all of which have shown state-of-the-art performance in similar tasks [45, 46, 47]. During training, we employed the Adam optimizer [48] with an initial learning rate of 10^{-5}, which was reduced by a factor of 10 when the accuracy on the dev set failed to improve for 10 epochs.

3.2.3. Multimodal: Text and Image

For the multimodal experiments, we utilized the network architecture reported in [49, 50], where Vision Transformer (ViT) [51] was employed for image feature encoding, and multilingual BERT (mBERT) was used for the textual representation. To combine both modalities, we utilized BLOCK fusion [52], a multimodal fusion technique based on block-superdiagonal tensor decomposition [51].

As OCR text was available in the English dataset, we trained a model by merging embedding of tweet text with image embedding , and then applied BLOCK fusion for multimodal integration with OCR text embedding.

Due to the highly imbalanced nature of the dataset, we conducted an additional experiment to address this issue by applying augmentation techniques to the low minority class, aiming to create a balanced training set. For this purpose, we employed synonym augmentation using WordNet, which is available in the NLPAug data augmentation package.[5]

Evaluation measures: The official evaluation metric for the shared task is the F_1 score for the positive class. However, in our experiments, we expanded our analysis to include additional evaluation metrics such as overall accuracy, precision, recall, and F_1 scores, specifically focusing on the positive class.

4. Results and Discussion

Table 2 presents the results of the submitted systems and additional unimodal and multimodal experiments. In many cases, we achieved better results than the baseline for both the English and Arabic datasets when using different modalities. Notably, the text-only modality consistently outperformed other modalities. The performance of the image modality, on the other hand, was relatively poor. Among the different image-only models, the EfficientNet models showed relatively better performance. While our additional multimodal experiments demonstrated improved results compared to the baseline and the submitted systems, they still performed lower than the text-only modality. Further experiments are necessary to fully understand the limitations of the multimodal models.

[5] https://github.com/makcedward/nlpaug

5. Conclusion

In this paper, we present our experiments and findings on check-worthiness classification as part of the `CheckThat! Lab` shared task. We provide a comparative analysis of different modalities and report that the text-only modality yields better results overall. Through our experiments, we observe that data augmentation plays a crucial role in improving performance. Moving forward, our future plans involve the further development of multilingual and multimodal models capable of capturing information from diverse modalities and languages.

6. Acknowledgments

The work of F. Alam is partially supported by NPRP 13S-0206-200281 and NPRP 14C-0916-210015 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

References

[1] J. Lawrence, C. Reed, Argument mining: A survey, Computational Linguistics 45 (2020) 765–818.

[2] P. Nakov, D. Corney, M. Hasanain, F. Alam, T. Elsayed, A. Barrón-Cedeño, P. Papotti, S. Shaar, G. D. S. Martino, Automated fact-checking for assisting human fact-checkers (2021).

[3] F. Alam, S. Shaar, F. Dalvi, H. Sajjad, A. Nikolov, H. Mubarak, G. D. S. Martino, A. Abdelali, N. Durrani, K. Darwish, A. Al-Homaid, W. Zaghouani, T. Caselli, G. Danoe, F. Stolk, B. Bruntink, P. Nakov, Fighting the COVID-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society, in: Findings of EMNLP 2021, 2021, pp. 611–649.

[4] F. Alam, S. Shaar, F. Dalvi, H. Sajjad, A. Nikolov, H. Mubarak, G. Da San Martino, A. Abdelali, N. Durrani, K. Darwish, A. Al-Homaid, W. Zaghouani, T. Caselli, G. Danoe, F. Stolk, B. Bruntink, P. Nakov, Fighting the COVID-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society, in: Findings of the Association for Computational Linguistics: EMNLP 2021, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 611–649.

[5] N. Hassan, G. Zhang, F. Arslan, J. Caraballo, D. Jimenez, S. Gawsane, S. Hasan, M. Joseph, A. Kulkarni, A. K. Nayak, et al., ClaimBuster: The first-ever end-to-end fact-checking system, Proceedings of the VLDB Endowment 10 (2017) 1945–1948.

[6] P. Gencheva, P. Nakov, L. Màrquez, A. Barrón-Cedeño, I. Koychev, A context-aware approach for detecting worth-checking claims in political debates, in: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, 2017, pp. 267–276.

[7] T. Elsayed, P. Nakov, A. Barrón-Cedeño, M. Hasanain, R. Suwaileh, G. Da San Martino, P. Atanasova, Overview of the CLEF-2019 CheckThat!: Automatic identification and

verification of claims, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, LNCS, 2019, pp. 301–321.
[8] S. Shaar, A. Nikolov, N. Babulkov, F. Alam, A. Barrón-Cedeño, T. Elsayed, M. Hasanain, R. Suwaileh, F. Haouari, G. Da San Martino, P. Nakov, Overview of CheckThat! 2020 English: Automatic identification and verification of claims in social media, CEUR Workshop Proceedings, 2020.
[9] P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, W. Mansour, B. Hamdan, Z. S. Ali, N. Babulkov, A. Nikolov, G. K. Shahi, J. M. Struß, T. Mandl, M. Kutlu, Y. S. Kartal, Overview of the CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news, LNCS (12880), 2021.
[10] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, The CLEF-2022 CheckThat! Lab on fighting the covid-19 infodemic and fake news detection, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, V. Setty (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2022, pp. 416–428.
[11] A. Barrón-Cedeño, F. Alam, T. Caselli, G. Da San Martino, T. Elsayed, A. Galassi, F. Haouari, F. Ruggeri, J. M. Struß, R. N. Nandi, G. S. Cheema, D. Azizov, P. Nakov, The CLEF-2023 CheckThat! Lab: Checkworthiness, subjectivity, political bias, factuality, and authority, in: J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, A. Caputo (Eds.), Advances in Information Retrieval, Springer Nature Switzerland, Cham, 2023, pp. 506–517.
[12] A. Barrón-Cedeño, F. Alam, A. Galassi, G. Da San Martino, P. Nakov, , T. Elsayed, D. Azizov, T. Caselli, G. Cheema, F. Haouari, M. Hasanain, M. Kutlu, C. Li, F. Ruggeri, J. M. Struß, W. Zaghouani, Overview of the CLEF–2023 CheckThat! Lab checkworthiness, subjectivity, political bias, factuality, and authority of news articles and their source, in: A. Arampatzis, E. Kanoulas, T. Tsikrika, S. Vrochidis, A. Giachanou, D. Li, M. Aliannejadi, M. Vlachos, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2023), 2023.
[13] F. Alam, A. Barrón-Cedeño, G. S. Cheema, S. Hakimov, M. Hasanain, C. Li, R. Míguez, H. Mubarak, G. K. Shahi, W. Zaghouani, P. Nakov, Overview of the CLEF-2023 CheckThat! lab task 1 on check-worthiness in multimodal and multigenre content, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.
[14] P. Atanasova, L. Màrquez, A. Barrón-Cedeño, T. Elsayed, R. Suwaileh, W. Zaghouani, S. Kyuchukov, G. Da San Martino, P. Nakov, Overview of the CLEF-2018 CheckThat! lab on automatic identification and verification of political claims, task 1: Check-worthiness, in: CLEF 2018 Working Notes. Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, 2018.
[15] P. Atanasova, P. Nakov, G. Karadzhov, M. Mohtarami, G. Da San Martino, Overview of the CLEF-2019 CheckThat! lab on automatic identification and verification of claims. Task 1: Check-worthiness, CEUR Workshop Proceedings, 2019.

[16] T. Elsayed, P. Nakov, A. Barrón-Cedeño, M. Hasanain, R. Suwaileh, G. Da San Martino, P. Atanasova, CheckThat! at CLEF 2019: Automatic identification and verification of claims, in: L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, D. Hiemstra (Eds.), Advances in Information Retrieval, ECIR '19, 2019, pp. 309–315.
[17] I. Jaradat, P. Gencheva, A. Barrón-Cedeño, L. Màrquez, P. Nakov, ClaimRank: Detecting check-worthy claims in Arabic and English, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, 2018, pp. 26–30.
[18] S. Vasileva, P. Atanasova, L. Màrquez, A. Barrón-Cedeño, P. Nakov, It takes nine to smell a rat: Neural multi-task learning for check-worthiness prediction, in: Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP '19, 2019, pp. 1229–1239.
[19] S. Shaar, F. Alam, G. Da San Martino, A. Nikolov, W. Zaghouani, P. Nakov, A. Feldman, Findings of the NLP4IF-2021 shared tasks on fighting the COVID-19 infodemic and censorship detection, in: Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, NLP4IF '21', Association for Computational Linguistics, Online, 2021, pp. 82–92.
[20] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, H. Mubarak, A. Nikolov, Y. S. Kartal, J. Beltrán, Overview of the CLEF-2022 CheckThat! lab task 1 on identifying relevant claims in tweets, in: N. Faggioli, Guglielmo andd Ferro, A. Hanbury, M. Potthast (Eds.), Working Notes of CLEF 2022—Conference and Labs of the Evaluation Forum, CLEF '2022, Bologna, Italy, 2022.
[21] S. Shaar, F. Haouari, W. Mansour, M. Hasanain, N. Babulkov, F. Alam, G. Da San Martino, T. Elsayed, P. Nakov, Overview of the CLEF-2021 CheckThat! lab task 2 on detecting previously fact-checked claims in tweets and political debates, in: [53], 2021.
[22] S. Shaar, N. Babulkov, G. Da San Martino, P. Nakov, That is a known lie: Detecting previously fact-checked claims, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL '20, 2020, pp. 3607–3618.
[23] S. Shaar, F. Alam, G. Da San Martino, P. Nakov, The role of context in detecting previously fact-checked claims, in: Findings of the Association for Computational Linguistics: NAACL 2022, 2022, pp. 1619–1631.
[24] I. Augenstein, C. Lioma, D. Wang, L. Chaves Lima, C. Hansen, C. Hansen, J. G. Simonsen, MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims, 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 2019, 2019, pp. 4685–4697.
[25] M. Mohtarami, R. Baly, J. Glass, P. Nakov, L. Màrquez, A. Moschitti, Automatic stance detection using end-to-end memory networks, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, 2018, pp. 767–776.
[26] M. Mohtarami, J. Glass, P. Nakov, Contrastive language adaptation for cross-lingual stance detection, 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 2019, 2019, pp. 4442–4452.

[27] P. Atanasova, P. Nakov, L. Màrquez, A. Barrón-Cedeño, G. Karadzhov, T. Mihaylova, M. Mohtarami, J. Glass, Automatic fact-checking using context and discourse information, Journal of Data and Information Quality (JDIQ) 11 (2019) 12.
[28] V. Nguyen, K. Sugiyama, P. Nakov, M. Kan, FANG: leveraging social context for fake news detection using graph representation, in: M. d'Aquin, S. Dietze, C. Hauff, E. Curry, P. Cudré-Mauroux (Eds.), CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020, 2020, pp. 1165–1174.
[29] J. Thorne, A. Vlachos, C. Christodoulopoulos, A. Mittal, FEVER: a large-scale dataset for fact extraction and VERification, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018, pp. 809–819.
[30] N. Hassan, C. Li, M. Tremayne, Detecting check-worthy factual claims in presidential debates, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15, 2015, pp. 1835–1838.
[31] A. Patwari, D. Goldwasser, S. Bagchi, TATHYA: A multi-classifier system for detecting check-worthy statements in political debates, in: E. Lim, M. Winslett, M. Sanderson, A. W. Fu, J. Sun, J. S. Culpepper, E. Lo, J. C. Ho, D. Donato, R. Agrawal, Y. Zheng, C. Castillo, A. Sun, V. S. Tseng, C. Li (Eds.), Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM, 2017, pp. 2259–2262.
[32] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, G. Da San Martino, A. Abdelali, H. Sajjad, K. Darwish, P. Nakov, Fighting the COVID-19 infodemic in social media: A holistic perspective and a call to arms, in: Proceedings of the International AAAI Conference on Web and Social Media, ICWSM '21, 2021, pp. 913–922.
[33] S. Shaar, F. Alam, G. Da San Martino, A. Nikolov, W. Zaghouani, P. Nakov, A. Feldman, Findings of the NLP4IF-2021 shared tasks on fighting the COVID-19 infodemic and censorship detection, in: Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, NLP4IF '21, 2021, pp. 82–92.
[34] R. Agez, C. Bosc, C. Lespagnol, J. Mothe, N. Petitcol, IRIT at CheckThat! 2018, in: [54], 2018.
[35] B. Ghanem, M. Montes-y Gómez, F. Rangel, P. Rosso, UPV-INAOE-Autoritas - Check That: Preliminary approach for checking worthiness of claims, in: [54], 2018.
[36] C. Hansen, C. Hansen, J. Simonsen, C. Lioma, The Copenhagen team participation in the check-worthiness task of the competition of automatic identification and verification of claims in political debates of the CLEF-2018 fact checking lab, in: [54], 2018.
[37] C. Zuo, A. Karakas, R. Banerjee, A hybrid recognition system for check-worthy claims using heuristics and supervised learning, in: [54], 2018.
[38] S. Shaar, M. Hasanain, B. Hamdan, Z. S. Ali, F. Haouari, A. Nikolov, M. Kutlu, Y. S. Kartal, F. Alam, G. Da San Martino, A. Barrón-Cedeño, R. Míguez, J. Beltrán, T. Elsayed, P. Nakov, Overview of the CLEF-2021 CheckThat! lab task 1 on check-worthiness estimation in tweets and political debates, in: [53], 2021.
[39] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, 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), Association for Computational

Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186.
[40] W. Antoun, F. Baly, H. Hajj, AraBERT: Transformer-based model for Arabic language understanding, in: Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, OSAC '20, Marseille, France, 2020, pp. 9–15.
[41] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems, 2014, pp. 3320–3328.
[42] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, an Diego, CA, USA, 2015.
[43] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[44] M. Tan, Q. V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, arXiv:1905.11946 (2019).
[45] F. Ofli, F. Alam, M. Imran, Analysis of social media data using multimodal deep learning for disaster response, in: Proceedings of the Information Systems for Crisis Response and Management, 2020.
[46] F. Alam, T. Alam, M. A. Hasan, A. Hasnat, M. Imran, F. Ofli, Medic: a multi-task learning dataset for disaster image classification, Neural Computing and Applications 35 (2023) 2609–2632.
[47] F. Alam, T. Alam, F. Ofli, M. Imran, Robust training of social media image classification models, IEEE Transactions on Computational Social Systems (2022).
[48] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: International Conference on Learning Representations, 2015.
[49] R. N. Nandi, F. Alam, P. Nakov, TeamX@DravidianLangTech-ACL2022: A comparative analysis for troll-based meme classification, in: Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 79–85.
[50] R. N. Nandi, F. Alam, P. Nakov, Detecting the role of an entity in harmful memes: Techniques and their limitations, in: Proceedings of the Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situations, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 43–54.
[51] A. Kolesnikov, A. Dosovitskiy, D. Weissenborn, G. Heigold, J. Uszkoreit, L. Beyer, M. Minderer, M. Dehghani, N. Houlsby, S. Gelly, T. Unterthiner, X. Zhai, An image is worth 16x16 words: Transformers for image recognition at scale, in: International Conference on Learning Representations, 2021.
[52] H. Ben-younes, R. Cadene, N. Thome, M. Cord, Block:bilinear superdiagonal fusion for visual question answeringand visual relationship detection, in: Proceedings of the 33st Conference on Artificial Intelligence (AAAI), volume 15, 2019. URL: https://ojs.aaai.org/index.php/AAAI/article/download/4818/4691.
[53] G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), CLEF 2021 Working Notes. Working Notes of CLEF 2021–Conference and Labs of the Evaluation Forum, 2021.
[54] L. Cappellato, N. Ferro, J.-Y. Nie, L. Soulier (Eds.), Working Notes of CLEF 2018–Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, 2018.