# FakeClaim: A Multiple Platform-driven Dataset for Identification of Fake News on 2023 Israel-Hamas War

Gautam Kishore Shahi[1] Amit Kumar Jaiswal[2] Thomas Mandl[3]

[1]University of Duisburg-Essen, Germany [2] University of Surrey, United Kingdom [3] University of Hildesheim, Germany

Code & Data

## Introduction, Goal and Contribution

**Introduction:** This work introduces the novel approach for the collection of fake claims during a crisis event like the 2023 Israel-Hamas war [1]. The approach is used to collect fake claims from YouTube, and a classification model is built to predict YouTube videos.

**Goal:** Our goal is to present an approach to collect and analyse the fake claims spread on social media during crisis events. On-time claim analysis can stop disseminating fake news and its side effects.

| Feature | Value |
|---|---|
| ClaimId | factly_6 |
| Claim | Visuals of Israelis protesting against Prime Minister Benjamin Netanyahu for war against Palestine. |
| Label | False |
| Claim URL | https://tinyurl.com/25vm9h2e |
| Article Title | March 2023 Video Shared as Recent Israelis Protest Against Netanyahu Amid Ongoing Israel-Hamas Conflict |
| Article text | Amid the ongoing Israel-Hamas conflict that began in October 2023, a video depicting a massive crowd .... (refer to above link) |
| Published date | October 22, 2023 |
| Language | English |
| Claim Source | YouTube |
| Link to Claim | https://www.youtube.com/watch?v=XDJP50w3ri |
| Claim(video) Data | Details extracted from claim(video)* |

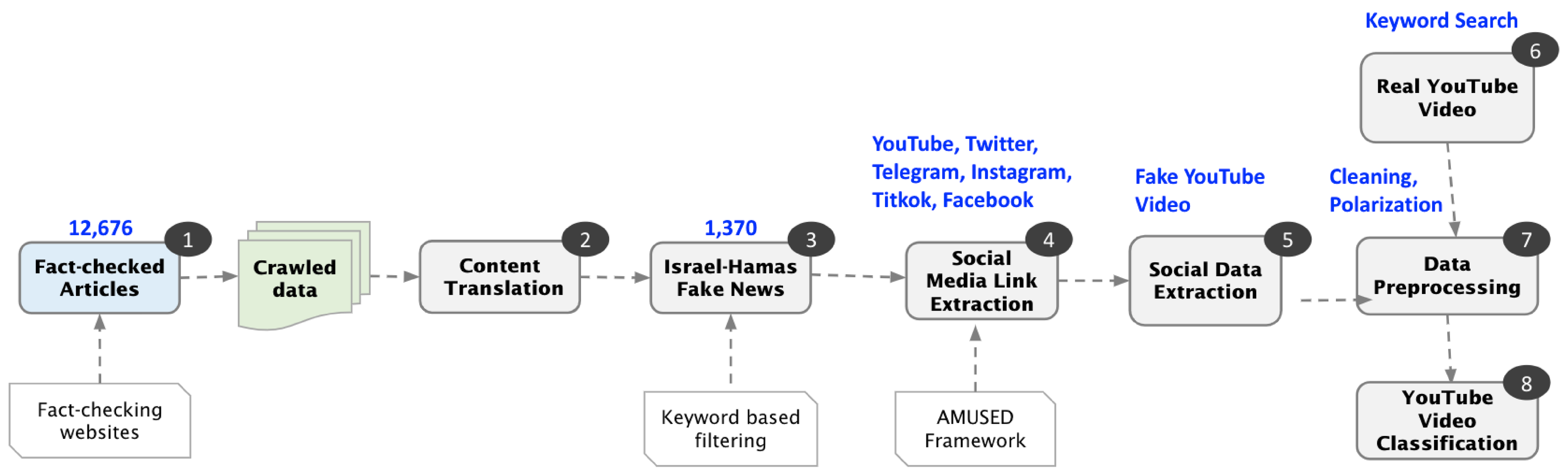
(a) A debunked claim with YouTube video link extraction from fact-checked articles


(b) Fake YouTube Video spread on 2023 Israel-Hamas war [2]

**Contributions:**

- We introduce the first fact-checked collection of 1,499 claims related to the 2023 Israel-Hamas war called FakeClaim.
- We demonstrate the utility of the dataset by forming a YouTube video classification by training state-of-the-art prediction models and find that user engagement, comments on the video, evidence pages and metadata significantly contribute to model performance.
- Finally, we benchmark our data with a fine-tuned version of the pre-trained embedding model, namely, the Universal sentence encoder [3], which jointly ranks evidence pages and performs veracity prediction.

## Data Collection & Preprocessing




**Dataset Preprocessing** The approach for data collection and classification model is shown in the above figure. An example of a claim collected using an AMUSED framework [2] is shown in the above table (a). After collection, articles are preprocessed by removing special characters, URLs, and emoji and applying a state-of-the-art classification model. After pre-processing, we computed frequent words and sentiment polarity for each comment as values in [-1,+1] and aggregated for each video. Overall, after the data cleaning and preprocessing steps, which excluded videos without any comments, the dataset consists of 756 unique videos and 166,645 comments, with 301 videos of the fake category with 85,228 comments and 455 of the true category with 81,417 comments.

## Problem Settings & Experiments

**Problem Formulation:** Let D be a fake news detection dataset containing N samples. In the social media setting, we define the dataset as

$$\mathcal{D} = \{\mathcal{V}, \mathcal{U}, \mathcal{C}, \mathcal{F}\}$$

where $\mathcal{V} = \{v_1, v_2, v_3, ...., v_n\}$ is a set of textual descriptions of **YouTube videos**, $\mathcal{U} = \{u_1, u_2, u_3, ...., u_n\}$ is a set of related **social users** engaged in conversation related to the corresponding YouTube videos. $\mathcal{C}$ represent **user's comments** on videos, in which $c \in \mathcal{C}$ and $\mathcal{F}$ depicts the claim and provides correct background information to debunk the claim which is extracted from **fact-checked article**, $f \in \mathcal{F}$, R represents the set of **social audience engagements and background truth** as the composition of $\mathcal{V}, \mathcal{U}, \mathcal{C}$ in which $r \in R$ is defined as a quadruplet $\{(v, u, c, f) | v \in V, u \in U, c \in C, f \in F\}$ (i.e. user u has given c comments on YouTube video V and f was background truth provided by fact-checkers). We formulate the problem as follows: Problem definition (Fake News Detection on YouTube Video): Given a YouTube video dataset D = V, U, C, F and ground-truth training labels $y_{train}$, the goal is to obtain a classifier f that, given test videos $V_{train}$, is able to predict the corresponding veracity labels $y_{test}$.

**Dataset:**

- The table shows the distribution of claims debunked from different social media platforms.
- Some fact-checked article debunks multiple claims from different social media platforms, so we got multiple claims from the same fact-checked articles.

| Claim Source | Number of Claims |
|---|---|
| Facebook | 579 |
| Twitter | 183 |
| YouTube | 389 |
| Tiktok | 186 |
| Instagram | 90 |
| Telegram | 72 |
| Total(*) | 1,499 |

**Results**

- We fed our model with concatenated pairs of video titles spanning comments and fact claims. We found out that the addition of fact-checked articles boosts the precision and F1 scores.
- On the contrary, we score the polarisation for each fact-checked article, where most articles and comments are highly skewed toward the words ' Israel,' 'Hamas,' and 'Gaza'. However, due to the skewness of the distribution of the sentiments, the comments tend to be high.
- Our model performance fluctuates in the range of 72% - 74%. Real videos are used without any fact-checked articles. Our current results are obtained without any additional polarisation features.

**Table 1.** Results with pre-trained word embedding models for classification of Fake claims. Model performance is based on the F1 score, best models are in bold.

| Model \ Class & Features | fake (0) | | | | | | real (1) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Video+ Comments+ Claims | | | Video+ Comments | | | Video+ Comments+ Claims | | | Video+ Comments | | |
| | Scores | | | | | | | | | | | |
| | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 |
| GN-Swivel-20D | 0.831 | 0.814 | 0.822 | 0.748 | 0.772 | 0.76 | 0.822 | 0.828 | 0.825 | 0.768 | 0.77 | 0.769 |
| Universal Sentence Encoder | 0.874 | 0.867 | **0.87** | 0.803 | 0.798 | **0.80** | 0.868 | 0.882 | **0.875** | 0.796 | 0.813 | **0.804** |
| RoBERTa | 0.69 | 0.75 | 0.72 | 0.66 | 0.72 | 0.69 | 0.64 | 0.69 | 0.664 | 0.70 | 0.69 | 0.689 |

**Conclusion & Future Work:**

We present a timely analysis of the 2023 Israel-Hamas war, providing a FakeClaim dataset that refers to the original fake posts and further scrapped YouTube videos and comments for the analysis.

- Our classification problem for fake video descriptions utilises textual information extracted from videos, comments, claims and evidence, where we found improvement in F1 scores resulting from the amalgamation of different features.
- The study can be useful in fighting fake videos and help to mitigate fake news.
- The limitations include the issue that the true class was not manually checked or fact-checked. We intend to obtain labels for correct videos as part of our future work.
- Finding around 1,499 claims in three weeks of conflict shows that the volume of claims is high, and an automated approach can help to debunk fake posts.
- The claims are spread over multiple platforms like YouTube, Twitter and Facebook. In future work, we intend to collect, explore and investigate FakeClaim data, given the diversity of online comments from social platforms such as TikTok.

**References:**

[1] Israel–Hamas war Wikipedia page, https://en.wikipedia.org/wiki/2023_Israel-Hamas_war

[2] Shahi, G. K., & Majchrzak, T. A. (2021, October). Amused: an annotation framework of multimodal social media data. In International Conference on Intelligent Technologies and Applications (pp. 287-299). Cham: Springer International Publishing.

[3] Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., ... Kurzweil, R. (2018). Universal sentence encoder. arXiv preprint arXiv:1803.11175.

[4] YouTube Video, url:https://www.youtube.com/watch?v=Y_QZ87uo01Iab_channel=AlJazeeraEnglish