CAVES: A Dataset to Facilitate Explainable Classification and Summarization of Concerns towards COVID Vaccines

Soham Poddar[1*], Azlaan Mustafa Samad[1], Rajdeep Mukherjee[1], Niloy Ganguly[1,2], Saptarshi Ghosh[1]
[1] Indian Institute of Technology, Kharargpur
[2] Leibniz University of Hannover

SIGIR 2022
Madrid

CAVES - "Concerns About Vaccines with Explanations and Summaries"

- 9,921 Anti-Vax tweets labelled with concerns about vaccines in a Multi-Label setting.
- Separate tweet excerpt as explanations for each label.
- Summaries of tweets in each class.

Anti-vax concerns (Labels)

- Vaccines are **Unnecessary**
- Vaccines should not be **Mandatory**
- Against Big **Pharma**
- Against **Political** agendas
- Deeper **Conspiracy** theories
- **Country** of origin of vaccine
- Vaccines have been **Rushed**
- **Ingredients** of vaccines
- **Side-effects** of vaccines
- Vaccines are **Ineffective**
- **Religious** concerns
- No specific concern

1. Social Importance

- Prior datasets –
  ◘ Identify broad level vaccine-stance from tweets *(Anti-Vax, Pro-Vax or Neutral)*
  ◘ No datasets for automatically identifying specific anti-vaccine concerns of people.
- Important to identify specific anti-vaccine concerns for suitable counter-arguments to people.
- Summarizing tweets can help authorities gain insights about nuances of concerns at a given time and place.

2. Importance in NLP

- Explainable multi-label problem – First dataset to contain separate explanations for each of the multiple labels.
- Multi-label classification is challenging: *Vocabulary used in tweets of different classes are similar and overlapping.*
- Summaries for each class can be used for performing multi-document or tweet-stream summarization.

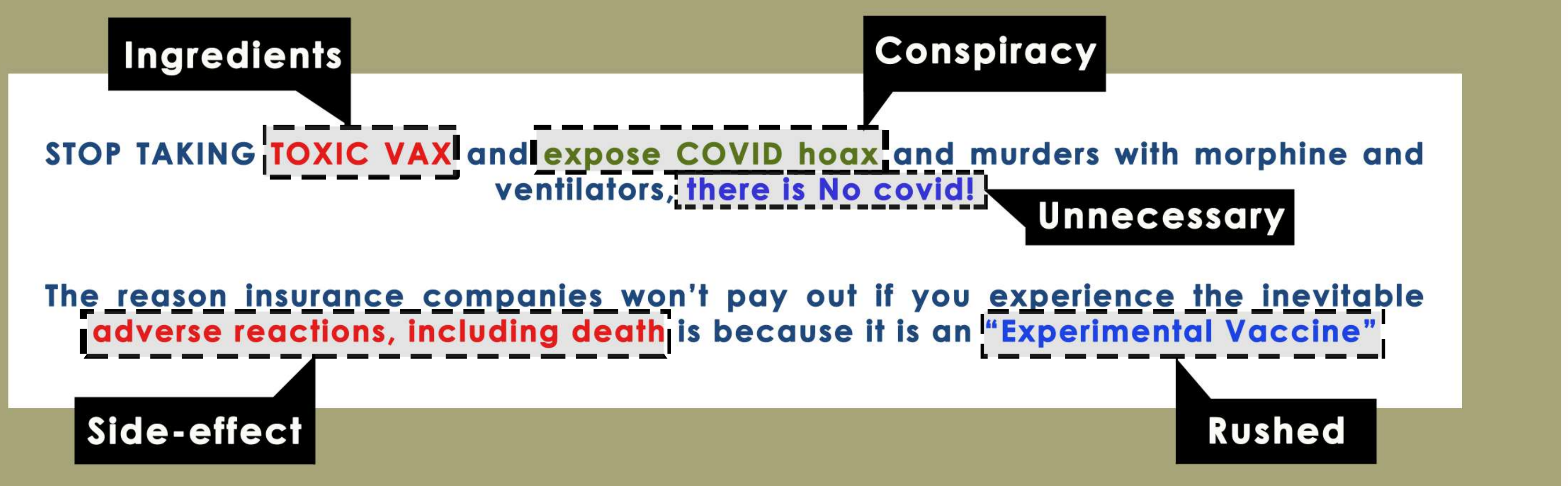
3. Gathering Tweets

- 100M vaccine related tweets collected between Jan 2020 - Oct 2021 using vaccine related keywords. *(E.g., "vaccine", "moderna", "covishield")*
- Only 10-15% are Anti-Vax, cannot take random sample!
- Automated filtering of Anti-Vax tweets:
  ◘ CT-BERT based vaccine stance classifier from our previous work. ◘ Retained tweets predicted Anti-Vax with high confidence (<80%)
- **Random sample of 11k tweets taken for annotation.**

4. Dataset preparation- Labels and Explanations

- 3 annotators labeled 11k tweets into 12 classes and marked corresponding tweet excerpts as explanations.
- A class was assigned to a tweet if marked by ≥ 2 annotators *(Krippendorff's α= 0.9557)*
- Union of individual explanations for a particular class in a particular tweet were taken *(given that class was assigned to the tweet)*

Explainable Multi-Label Example:



Ingredients | Conspiracy

STOP TAKING **TOXIC VAX** and **expose COVID hoax** and murders with morphine and ventilators, **there is No covid!**

Unnecessary

The reason insurance companies won't pay out if you experience the inevitable **adverse reactions, including death** is because it is an **"Experimental Vaccine"**

Side-effect | Rushed

5. Dataset preparation- Summaries

- 3 separate annotators summarized the tweets in each class in 200-250 words.
- Quality validated using "prolific" crowdsourced workers: Summaries had average score > 4 on all metrics- *Consistency, Fluency and Relevance.*

Summary Example:

Pharma

They believe that these companies are hiding data indicate harmful side-effects and rushing the vaccine process to make large amounts of money for themselves and shareholders. The tweets also indicate that FDA, NIH, Fauci, Boris Johnson, WHO, CDC, are all in on this plan to knowingly push a dangerous product because it will make them all rich ...

Mandatory

There is worry that there will be vaccine mandates that will escalate out of control such as kids not being able to go to school without a covid-19 virus. It is believed that vaccine mandates are completely unethical, unconstitutional, and some are even comparing mandatory vaccine passports to events that occured during the nazi germany period ...

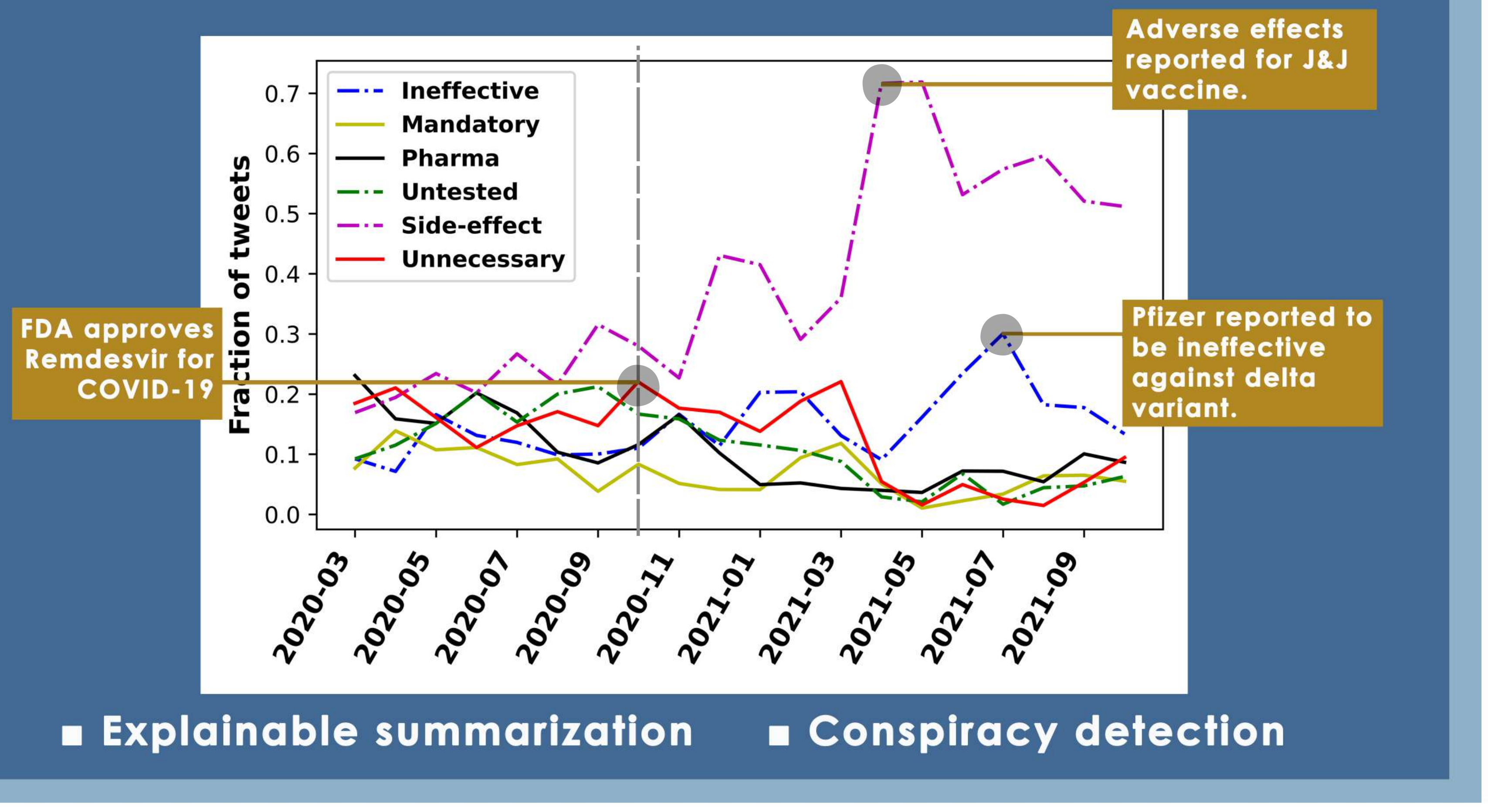
6. Tasks on the dataset

- Multi-Label Classification
- Explainable Multi-Label Classification
- Summarization

(Benchmark scores for each are given in the paper.)

7. Other Potential Uses

- Distribution of concerns over time



FDA approves Remdesvir for COVID-19

Adverse effects reported for J&J vaccine.

Pfizer reported to be ineffective against delta variant.

- Explainable summarization
- Conspiracy detection

contact:
sohampoddar26@gmail.com

Get the Dataset:
github.com/sohampoddar26/caves-data

[*] Received SIGIR student travel grant