Argumentative Stance Prediction: An Exploratory Study on Multimodality and Few-Shot Learning

Arushi Sharma*, **Abhibha Gupta***, Maneesh Bilalpur*
School of Computing and Information, University of Pittsburgh

MOTIVATION

Does **multimodality** improve argumentative **stance prediction**?

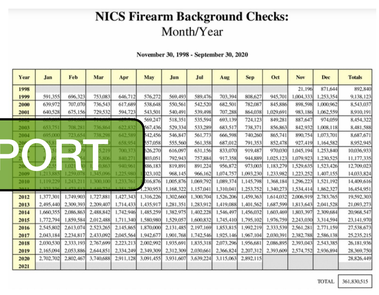
Tweets

Images

Gun control

The past year has seen more #NICS checks than any same month. #Gunsense #guncontrol is dead.

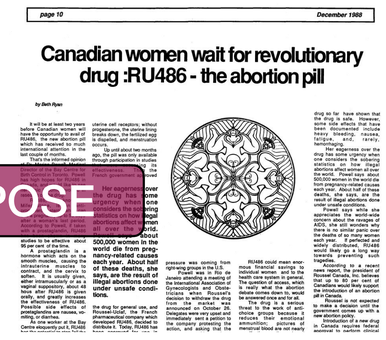
SUPPORT



Abortion

Turns out that women had to wait much, much longer - until 2017 - to get access to medical abortions. #riseuparchive #womenshistory #feminism #abortion.

OPPOSE



Importance

- Enhances understanding of **public opinion**, social dynamics and policy efficacy.
- **Immediate feedback** for policy makers.

Challenges

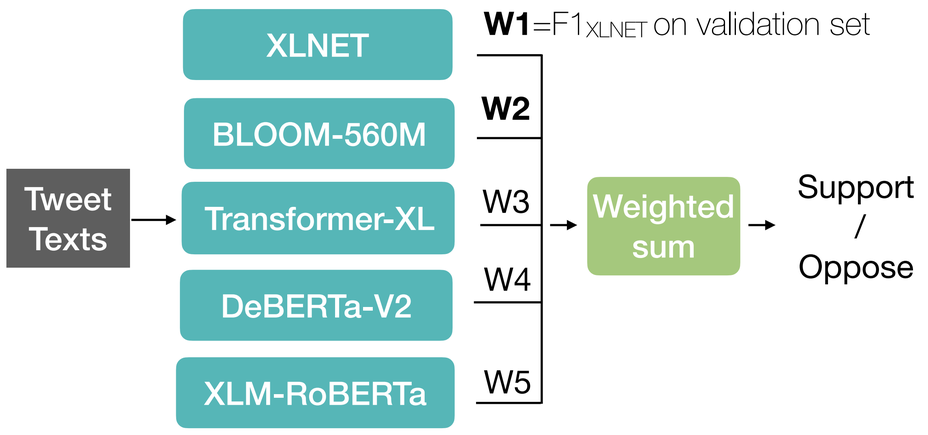
- **Lack of information** in brief tweets.
- **Effective representation** of text and image data in prediction models.

DATASET

- The abortion dataset is imbalanced by a 1:3 support:oppose stance ratio.
- Addressed through weighted cross-entropy loss, with higher weight for minority category.

EXPERIMENT

Text-based Transformer Models

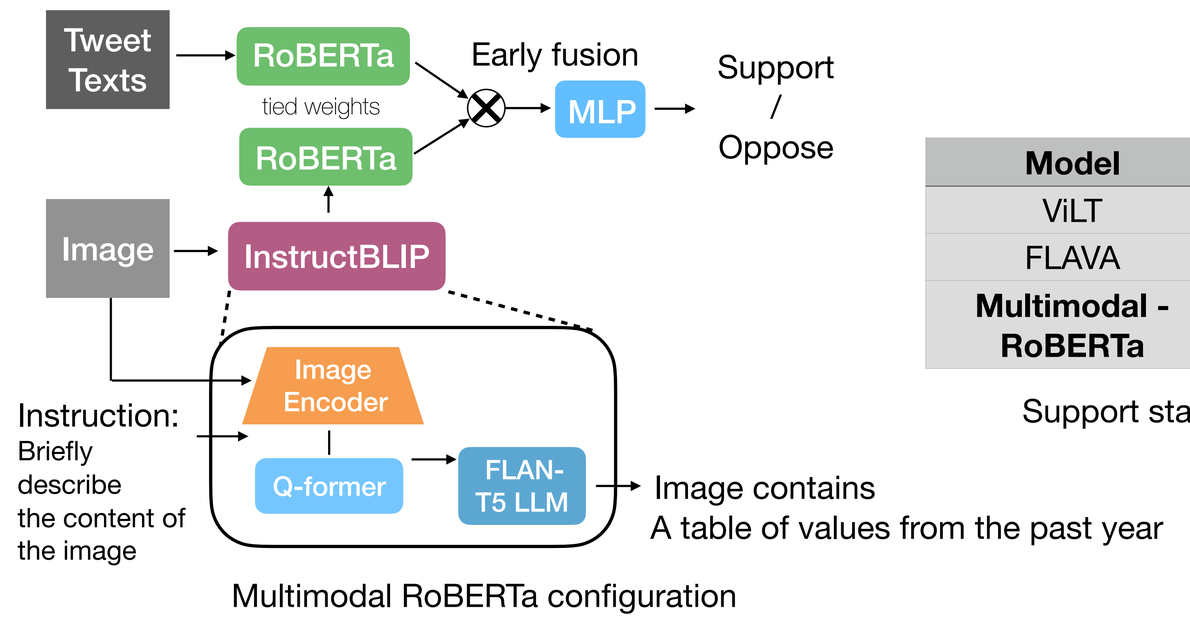


XLNet and BLOOM-560M received the predominant weights.

Model	P	R	F1
XLNET	0.619	0.924	0.741
BLOOM-1B	0.760	0.660	0.710
BLOOM-560M	0.707	0.898	<u>0.791</u>
Transformer-XL	0.571	0.881	0.693
DeBERTa-V2	0.560	0.710	0.630
XLM-RoBERTa	0.650	0.880	0.750
Ensemble	**0.743**	**0.906**	**0.817**

Support stance performance

Multimodal Transformer Models

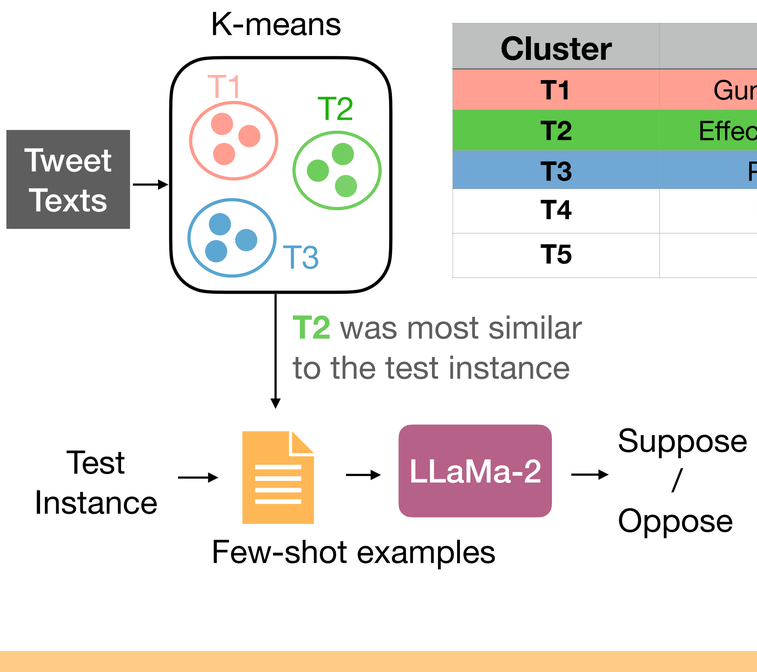


Multimodal RoBERTa configuration

Model	P	R	F1
ViLT	0.680	0.432	0.528
FLAVA	0.570	0.650	0.610
Multimodal - RoBERTa	**0.531**	**0.932**	**0.677**

Support stance performance

Few Shot Prompting Using LLM's



Cluster	Gun control	Abortion
T1	Gun violence as a mental health	Reproductive rights of women
T2	Effects of gun violence on children	#savethebabyhumans hashtag
T3	Pro-gun control politicians	Roe v Wade abortion case
T4	Racism and guns control	Religion and motherhood
T5	Trump and guns	Pro-life

Model	P	R	F1
Support-always	0.395	1.00	0.566
Zero-shot	0.440	0.290	0.350
Four-shot	0.420	0.640	0.500
Four-shot + K-means	**0.450**	**0.700**	**0.550**

Support stance performance

RESEARCH QUESTIONS

1. How well does text as a **stand-alone modality** perform?
2. Does **incorporating image** information improve performance?
3. How do Large-Language Models (LLMs) in **few-shot setting** compare against fine-tuned unimodal and multimodal models?

DISCUSSION

1. Ensemble method using text models performs the best (F1=0.817). BLOOM-560M, slightly underperforms the ensemble, but requires less computational resources.
2. Our configuration, Multimodal - RoBERTa, leverages InstructBLIP to convert image into text performs better than other models (ViLT, FLAVA).
3. Four-shot improves LLaMA-2 performance over zero-shot prompting. K-means clustering enhances context by providing similar examples to the test instance and improve inference.

FUTURE WORK

- Incorporating domain knowledge.
- Prompting methods such as Question Decomposition and Tree-of-Thought, offers a way to not only predict stance but also provide the rationale.