# Natural Language Processing for Rumour Verification in Social Media Conversations

Elena Kochkina

The Alan Turing Institute

WARWICK

# Presentation structure

Introduction/Motivation

Multitask learning for rumour verification

Estimating predictive uncertainty of rumour verification models

# Social media as a source of news

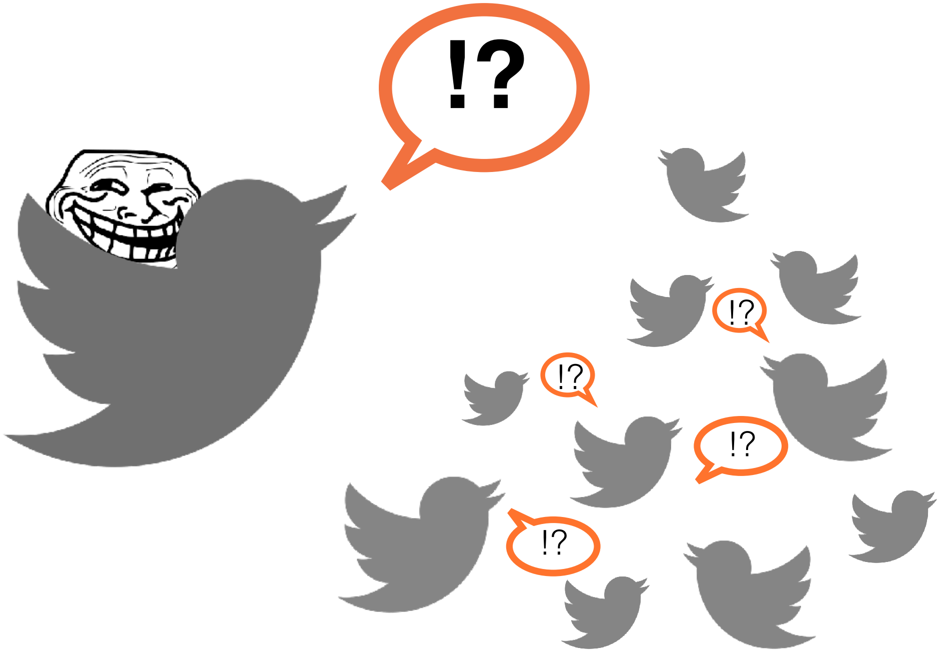
***67%** of American adults (ages 18+) were getting news from social media in 2017*

**Pew Research Center**

Absence of the verification on social media gives rise to the spread of misinformation.

Unverified claims can **spread rapidly** and reach **a wide audience**.

The **wide spread of false rumours carries a lot of risks**



# COVID-19 MISINFODEMIC

Poynter.

Media News | Training & Events | Support Us | About | Donate | Search

Fact-Checked by: Les Décodeurs
**2020/05/12 | France**
**MISLEADING: Japanese schools re-opened then were closed again due to a second wave of coronavirus.**

Read More

Fact-Checked by: Les Décodeurs
**2020/05/12 | France**
**FALSE: A Japanese Nobel Price said that SARS-CoV-2 was human made.**

Read More

Fact-Checked by: Les Décodeurs
**2020/05/12 | France**
**FALSE: Famous French blue cheese, roquefort, is a "medecine against Covid-19".**

Read More

Fact-Checked by: Les Décodeurs
**2020/05/12 | France**
**FALSE: Administrative documents French people need to fill to go out are a copy paste from 1940 documents.**

Read More

Fact-Checked by: Les Décodeurs
**2020/05/12 | France**
**FALSE: French Minister of Justice Nicole Belloubet threathened the famous anchor Jean-Pierre Pernaut after he criticized the government policy about the pandemic on air.**

Read More

Fact-Checked by: Les Décodeurs
**2020/05/12 | France**
**FALSE: French President Macron and its spouse are jetskiing during the lockdown.**

Read More

# Manual verification

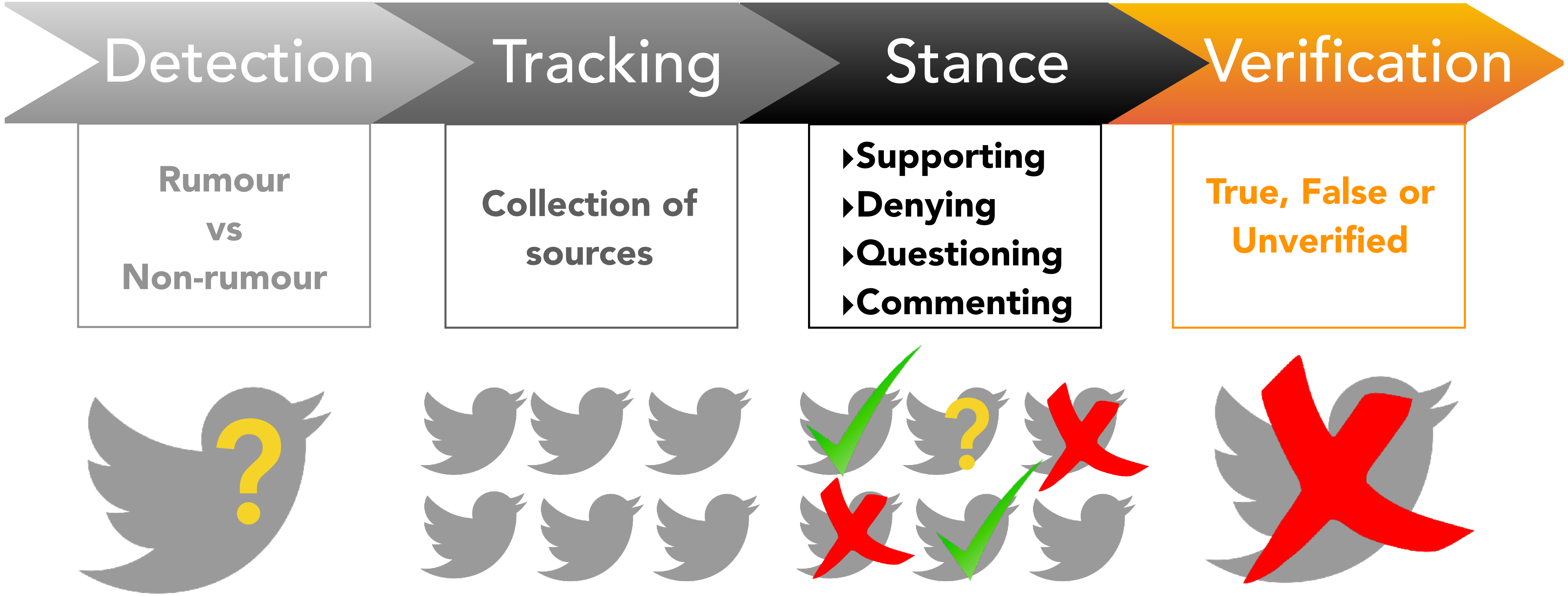
Traditionally rumour verification and fact-checking are performed **manually** by professionals


**Manual verification is not scalable** given the amount of mis-information being spread

Machine Learning and Natural Language Processing methods aim to assist with automating the verification process

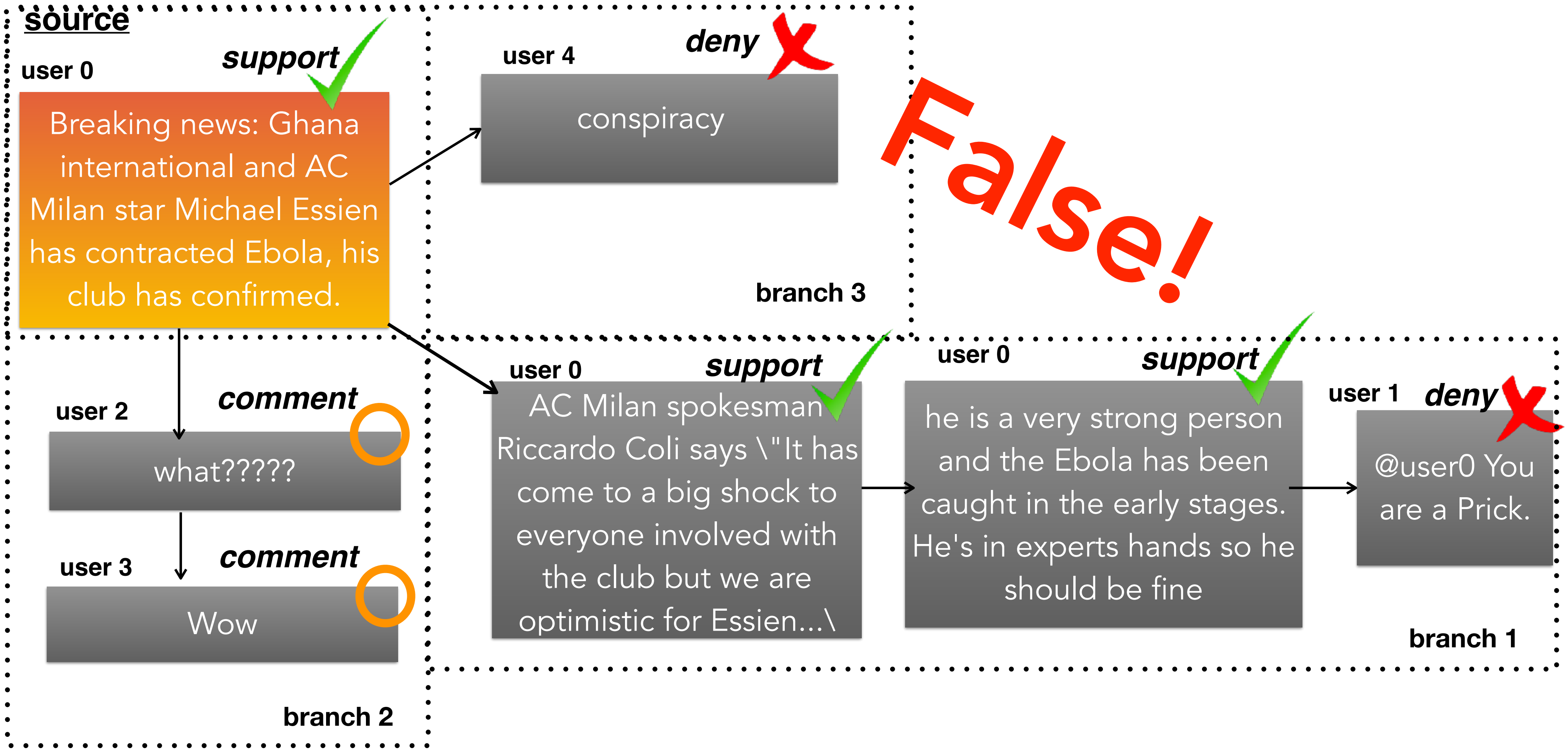
# Rumour verification pipeline

WARWICK




Zubiaga et al, 2017

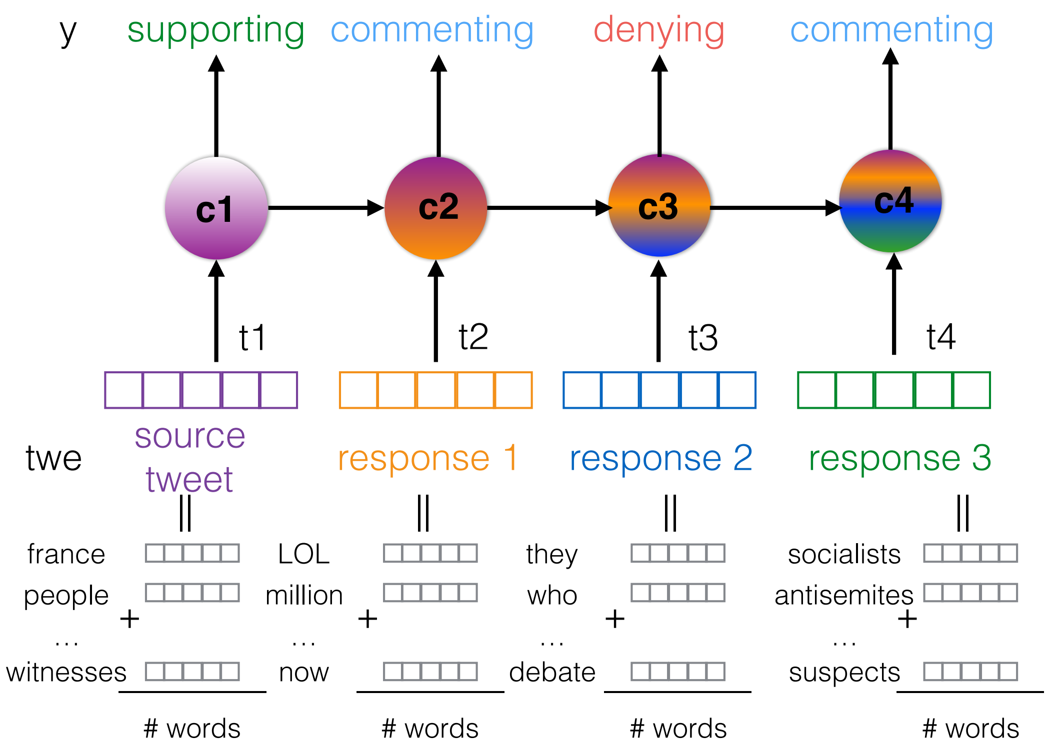
# Conversation example

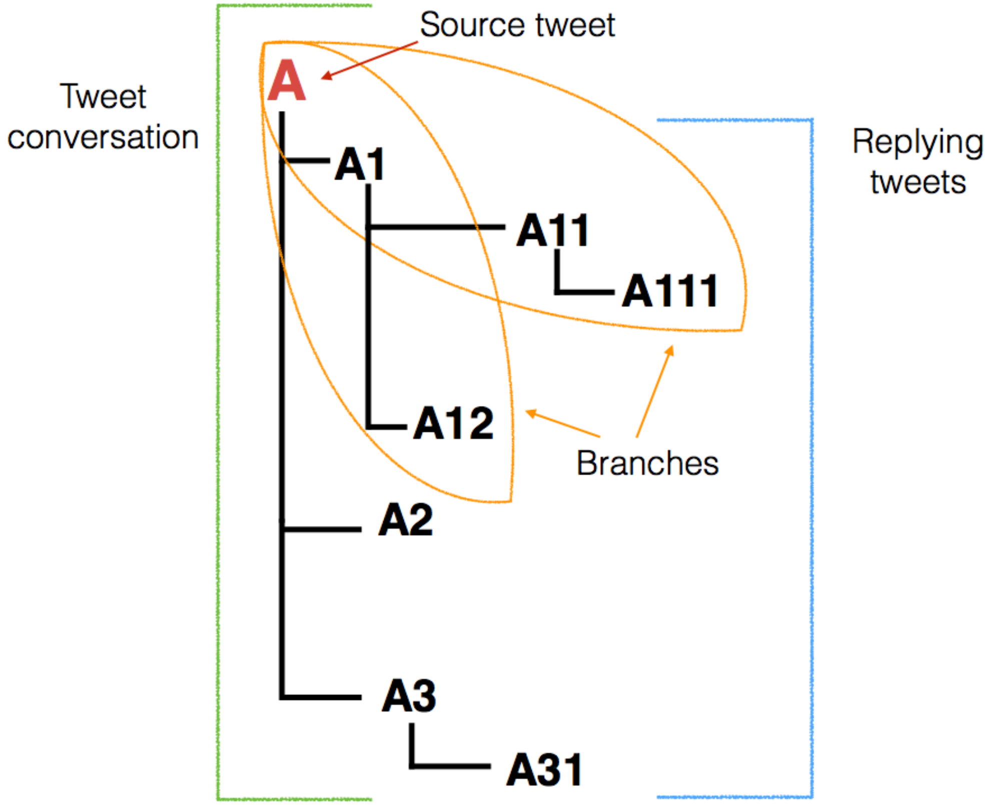


Rumours that attract a lot of scepticism are more likely to be proven false later

# Rumour stance classification

**Modeling conversation structure**
helps achieve better results
in stance classification comparing
to using individual tweets or tweet pairs



# Multi-task learning  for rumour verification

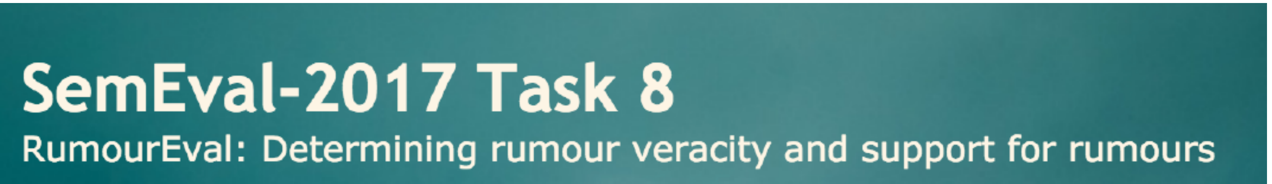
# RumourEval

The RumourEval shared task was proposed to **test the hypothesis regarding the synergy between stance and rumour veracity**.

RumourEval consists of 2 sub-tasks:

(A) Rumour stance classification and (B) Rumour veracity classification

RumourEval had 2 editions 2017 and 2019




# Veracity classification rumoureval-2017

| Team | Accuracy | RMSE |
| --- | --- | --- |
| DFKI DKT | 0.393 | 0.845 |
| ECNU | 0.464 | 0.736 |
| IITP | 0.286 | 0.807 |
| IKM | 0.536 | 0.763 |
| **NileTMRG** | **0.536** | **0.672** |

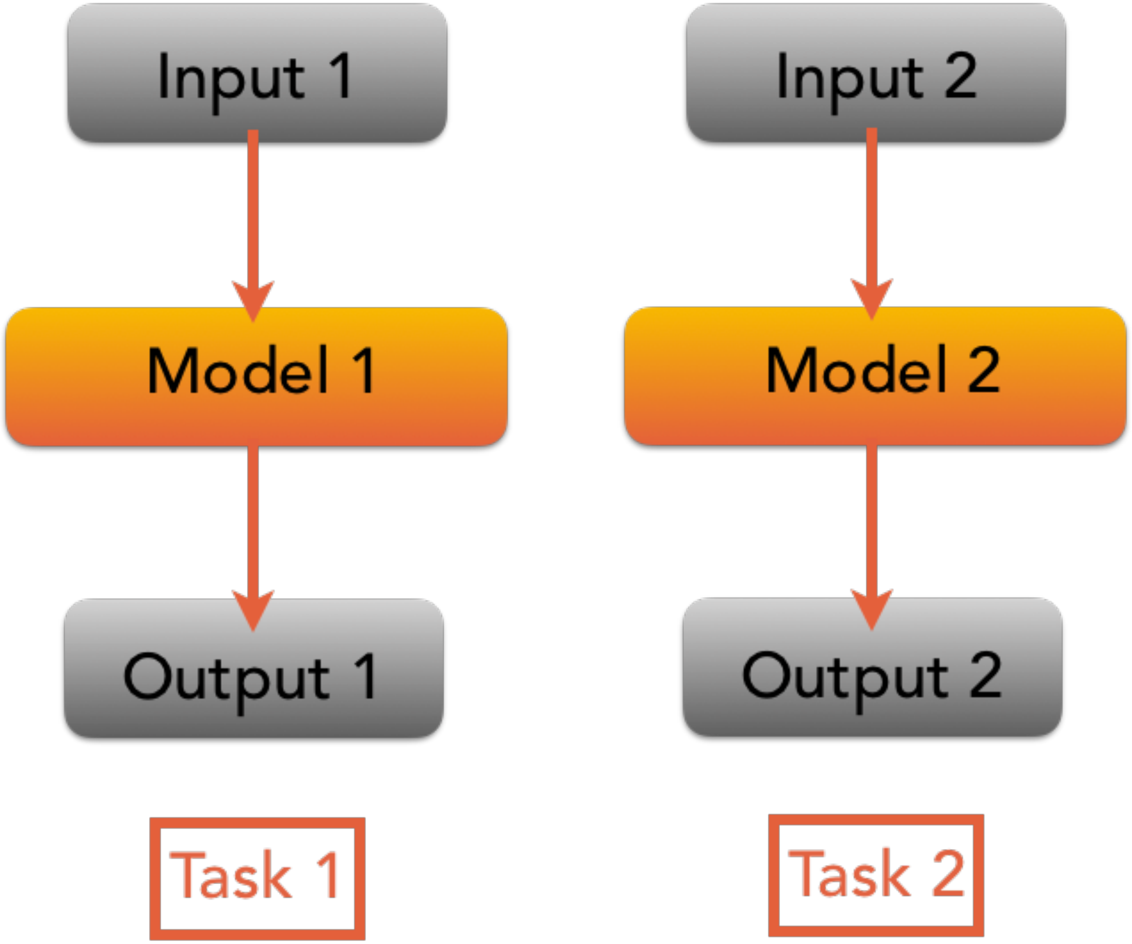
**Enayet and El-Beltagy (NileTMRG)**

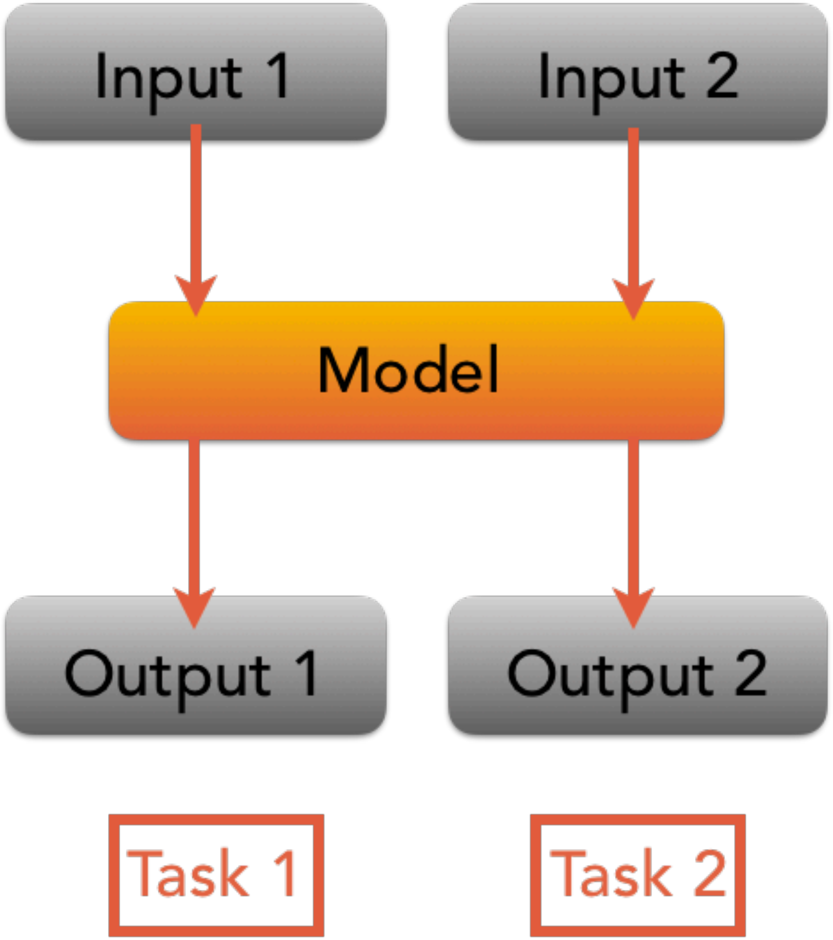
Linear **SVM** +

- bag-of-words (source tweet only)
- presence of URL
- presence of hashtag
- proportion of supporting, denying and querying tweets in the thread

# Multi-task learning





# Baselines

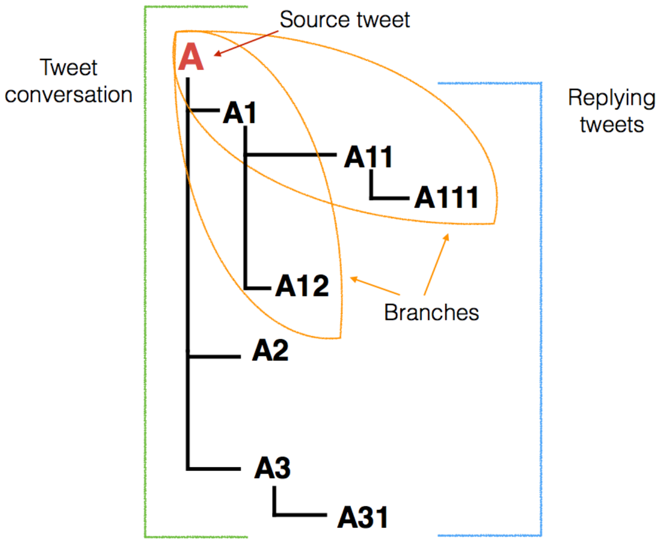
- **Majority** - always predicts majority class (True).

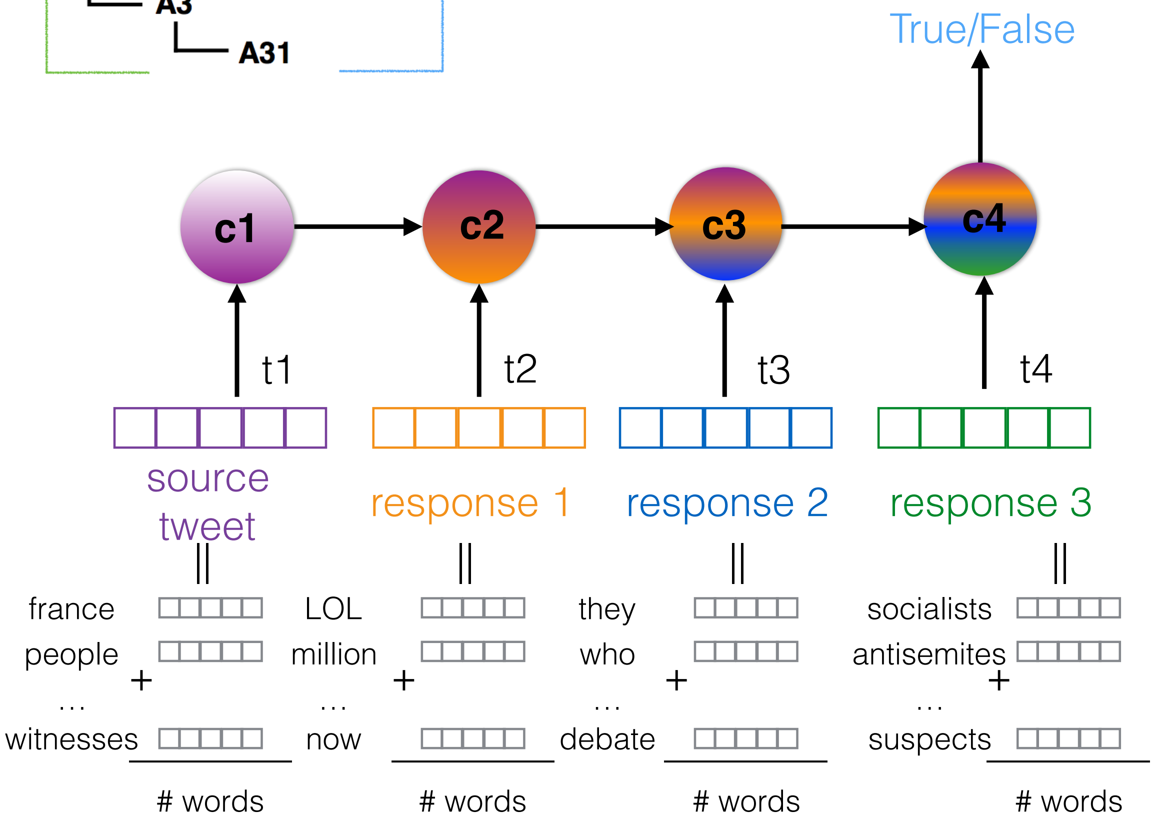
- **NileTMRG** (Enayet and El-Beltagy, 2017)

Linear SVM with BoW, presence of #, URL and **proportion of supporting, questioning and denying stances**

- **branchLSTM**

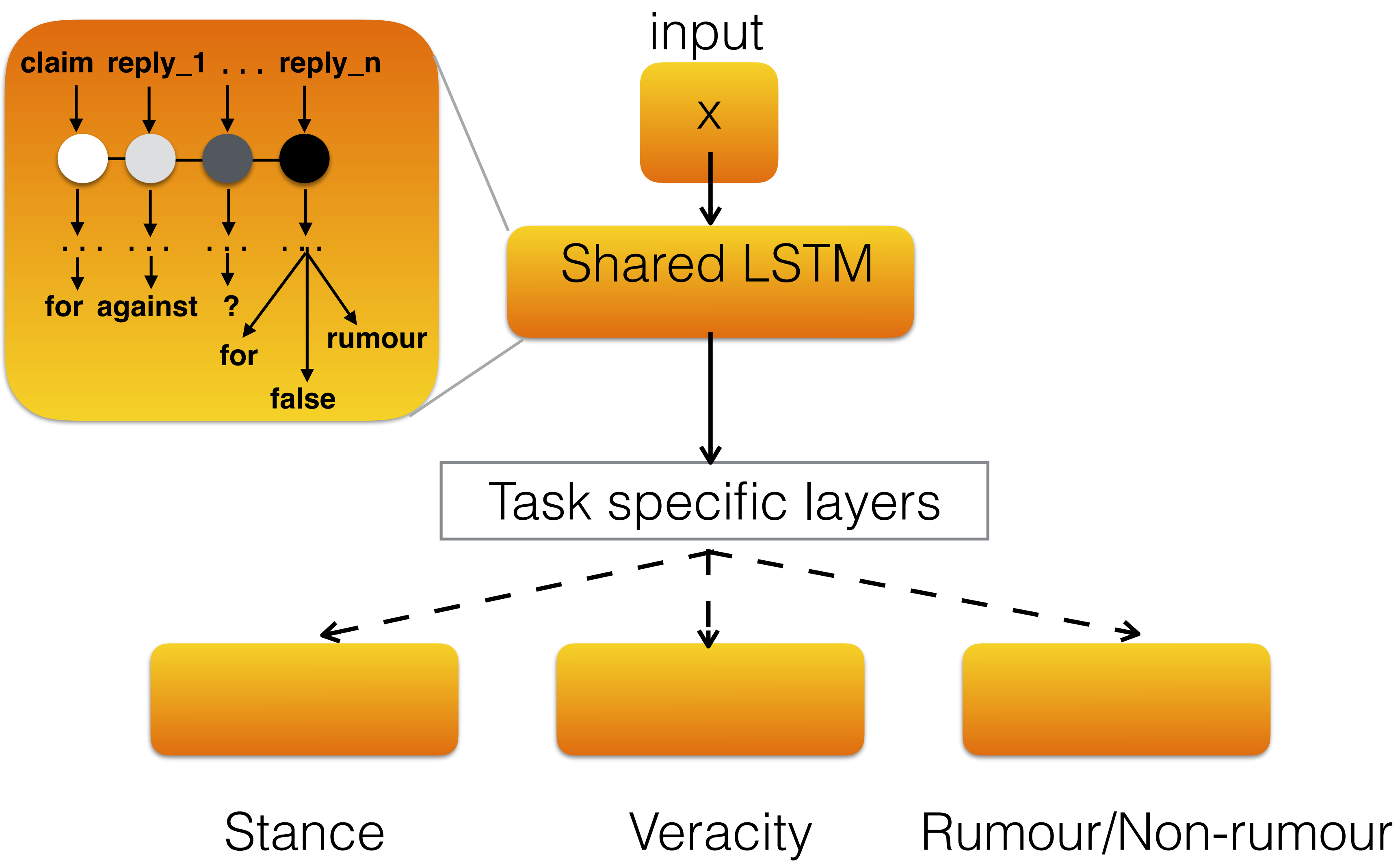
This is a single task model that utilises conversation structure. branchLSTM is a basis of MTL models.



# Multi-task models

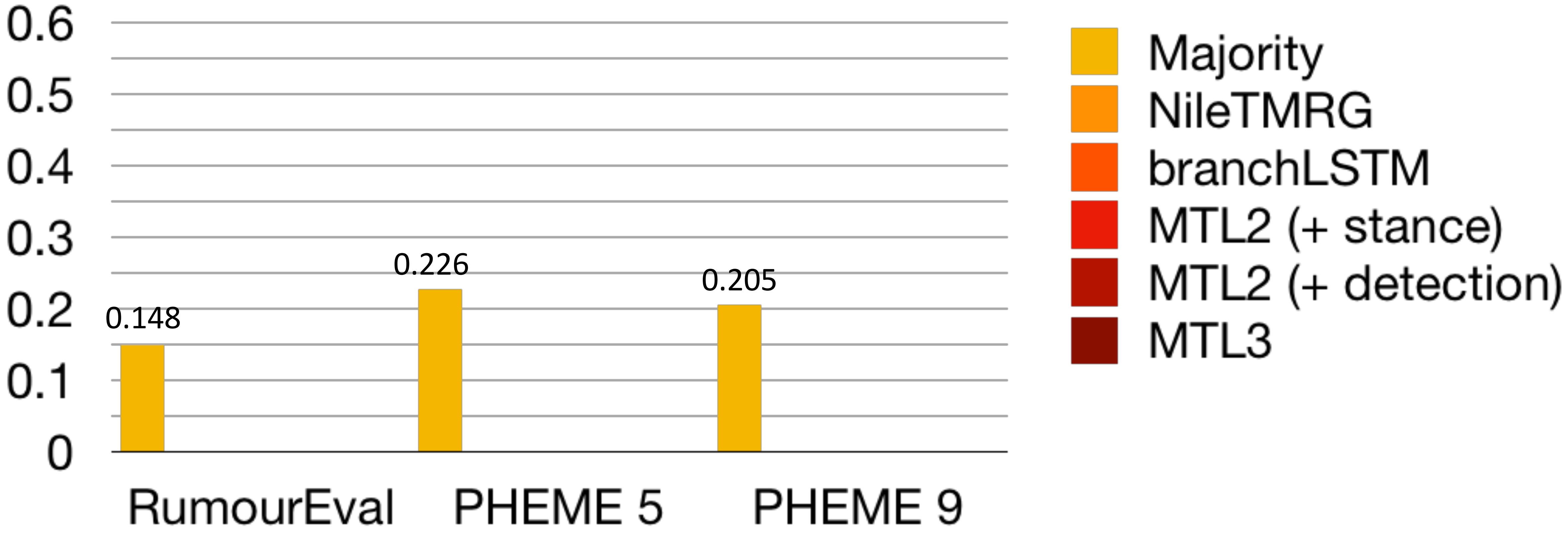
WARWICK



# Results

WARWICK

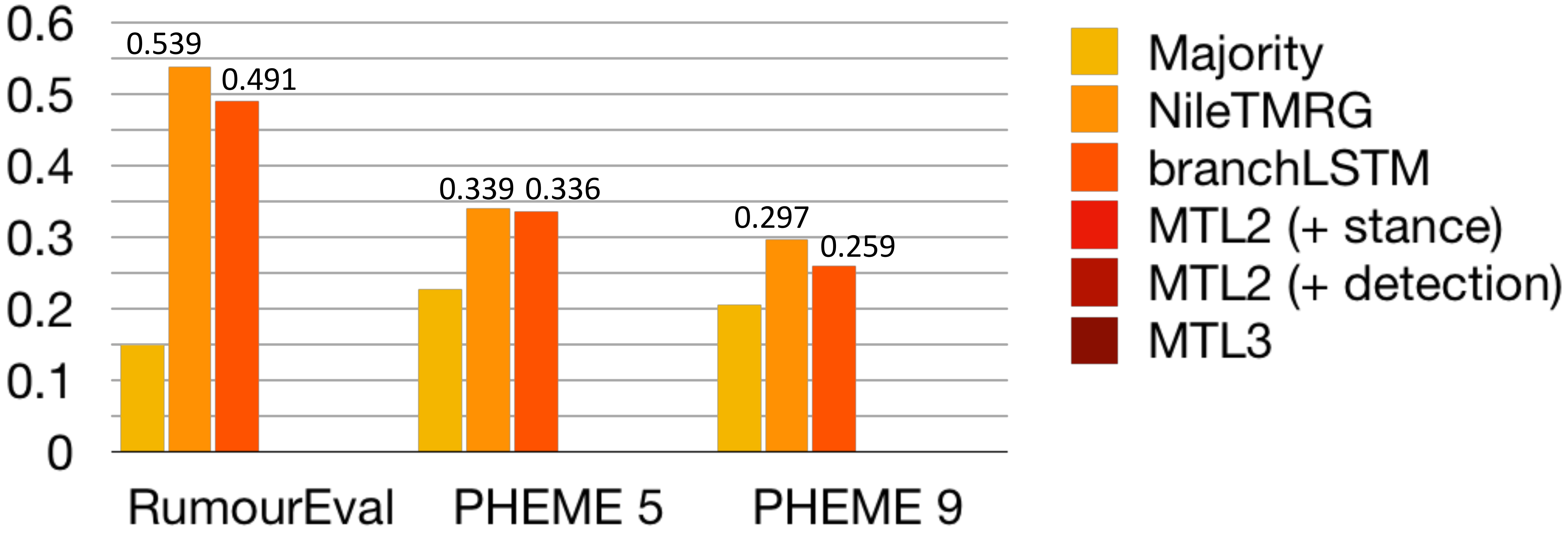
**Macro F1 score**



# Results

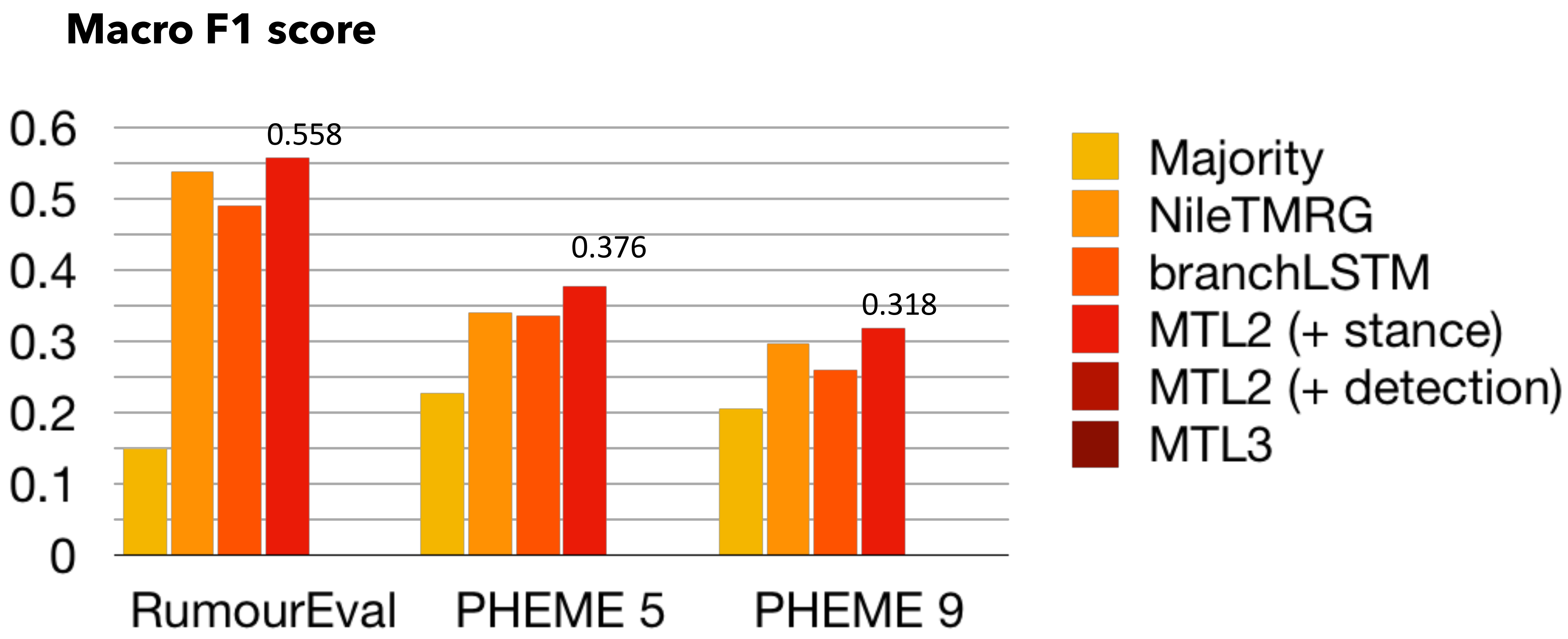
WARWICK

**Macro F1 score**



# Results

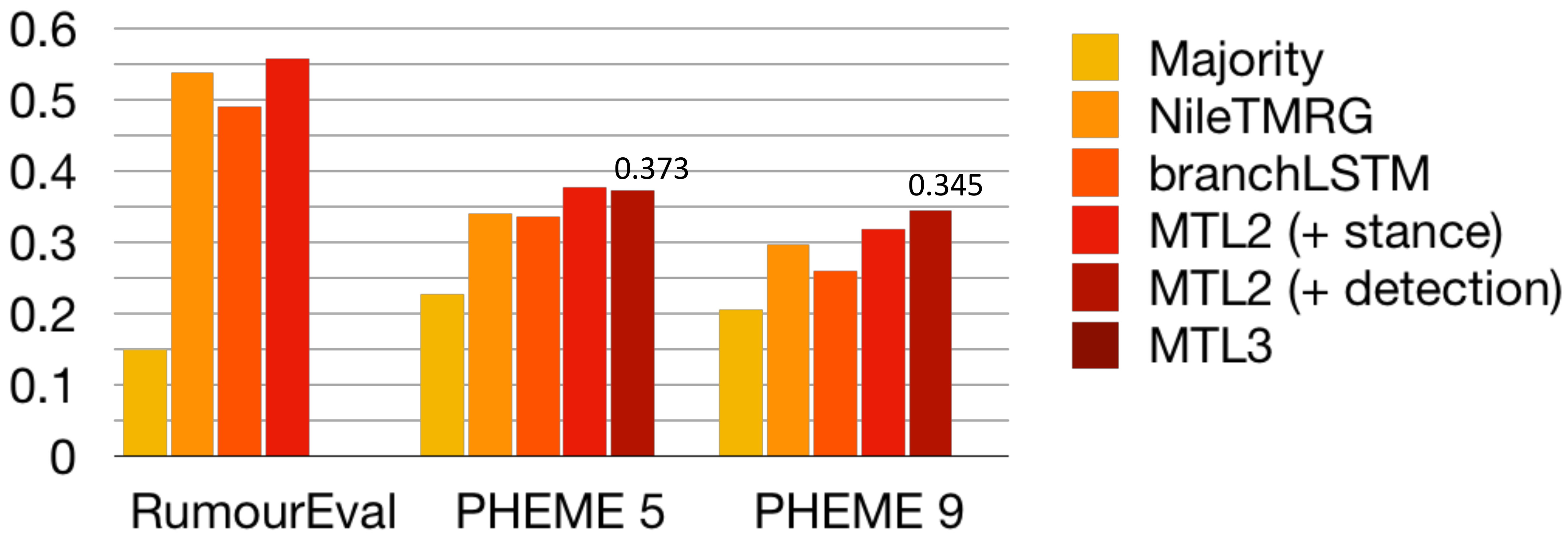



Macro F1 score
0.6
0.5
0.4
0.3
0.2
0.1
0
0.558
0.376
0.318
RumourEval
PHEME 5
PHEME 9
Majority
NileTMRG
branchLSTM
MTL2 (+ stance)
MTL2 (+ detection)
MTL3

# Results



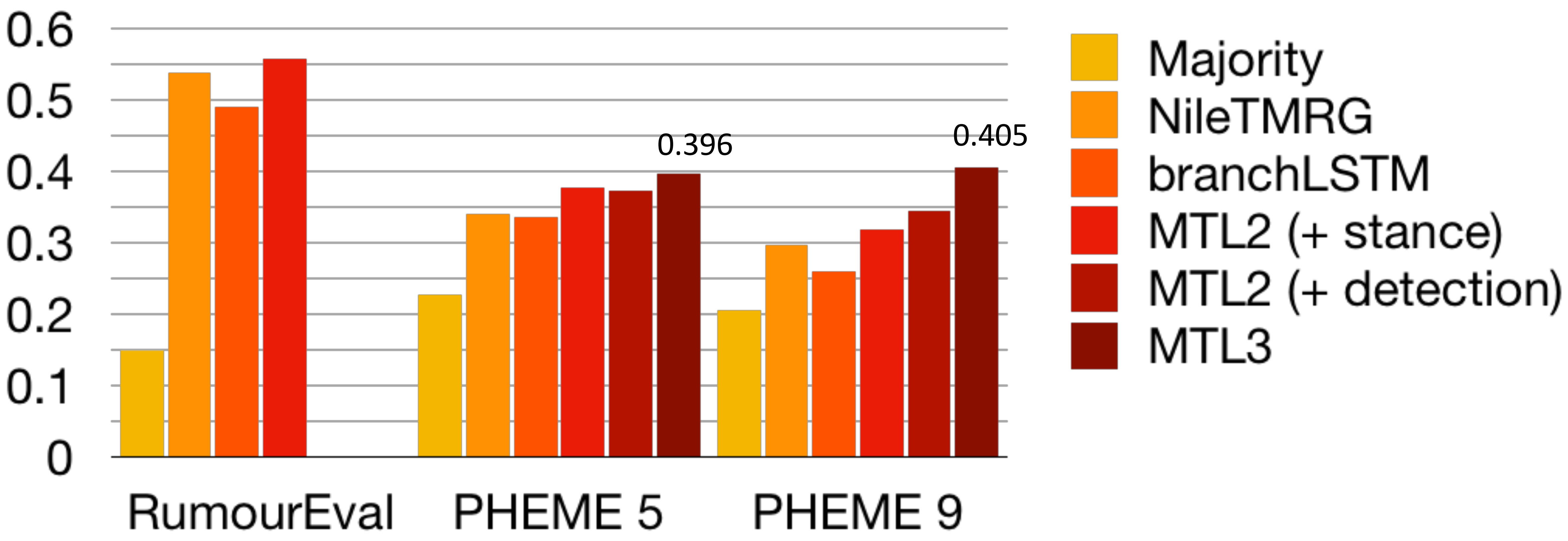
**Macro F1 score**



# Results

WARWICK

**Macro F1 score**



# Conclusions

WARWICK

- **Joint learning** of the tasks from the verification pipeline outperforms a single-learning approach to rumour verification.
- **Stance classification** contributes to improvements in rumour verification either as feature or as auxiliary task

# Estimating Predictive Uncertainty for Rumour Verification Models

# Estimating Predictive Uncertainty for Rumour Verification Models

WARWICK

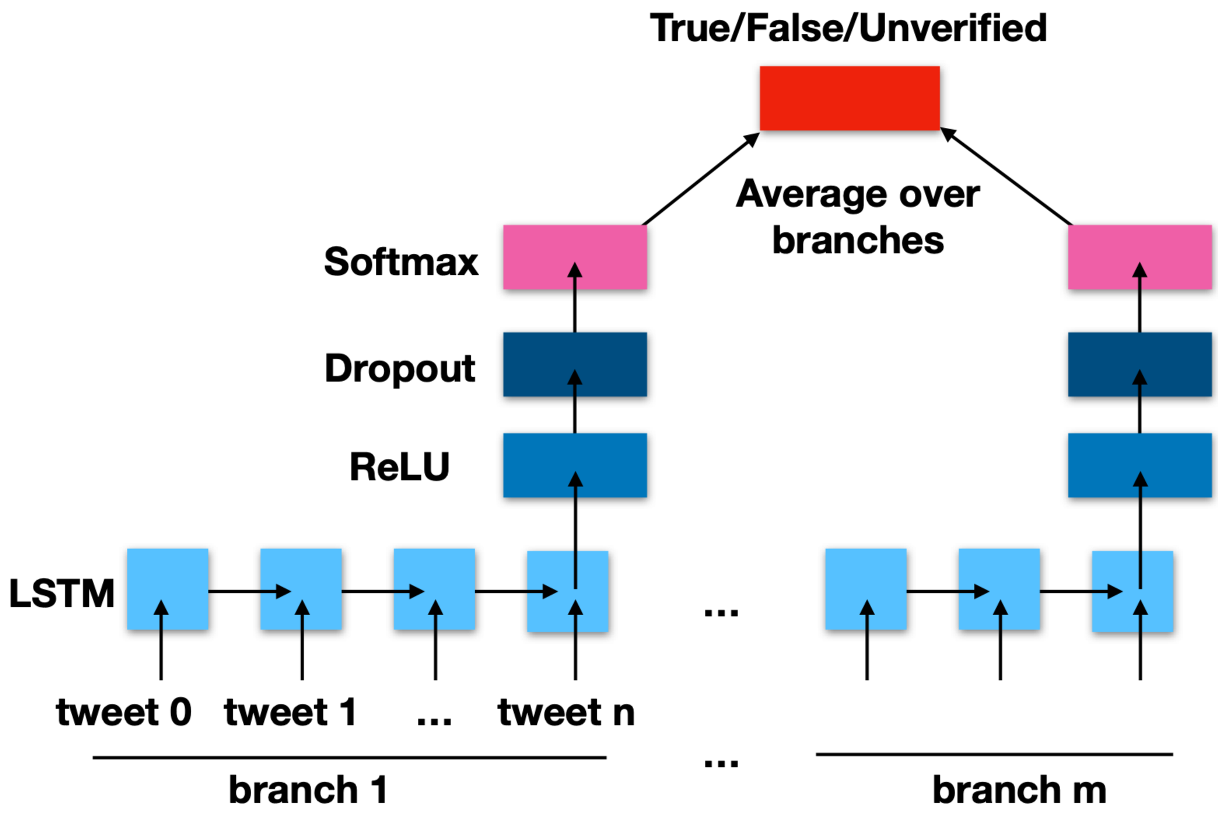
The cost of incorrect model predictions can be high, so we need to know when we should trust the model output

**We estimate predictive uncertainty** to gain understanding of a model's decisions and **filter out the 'hard' cases to be passed on to a human.**

The approach we adopt requires **minimal changes to a given model** and is relatively **computationally inexpensive**, thus making it **possible to apply to various architectures**

# Rumour Verification Model. Branch-LSTM

WARWICK




Input: $x_i, i \in [1, .., N]$- branch of tweets

$$u_i = f(x_i)$$

$$v_i = W_v u_i + b_v$$

$$p_i = softmax(v_i) = \frac{e^{v_i}}{\sum_{k=1}^{C} e^{v_i^k}}$$

Loss: $$l_1 = -\frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{C} y_i^k \log p_i^k$$

# Uncertainty

## Model Uncertainty

epistemic

Model uncertainty comes from **model parameters** and can be explained away given enough (i.e. an infinite amount of) data.

Method from (Gal and Gharamani, 2016)

## Data Uncertainty

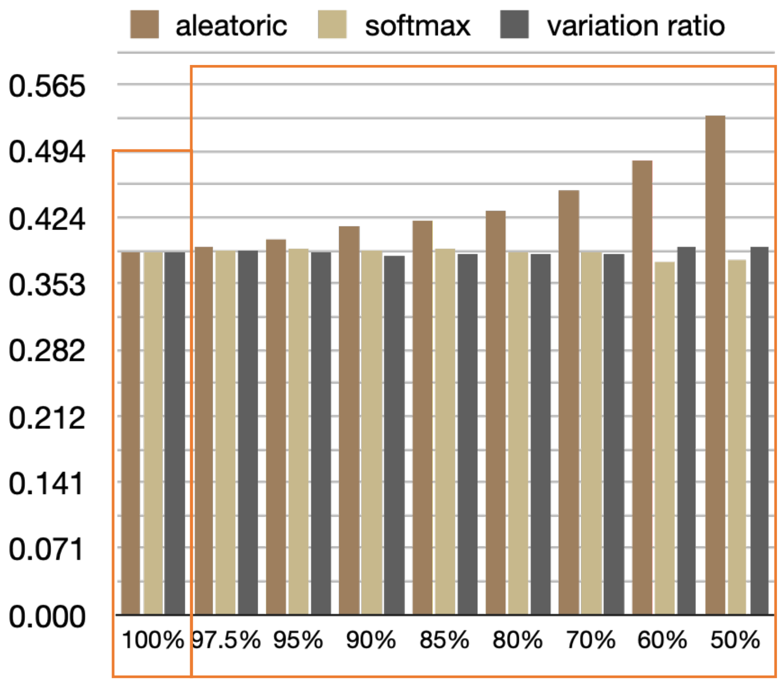
aleatoric

Data uncertainty is associated with **properties of the data**, such as imperfections in the measurements

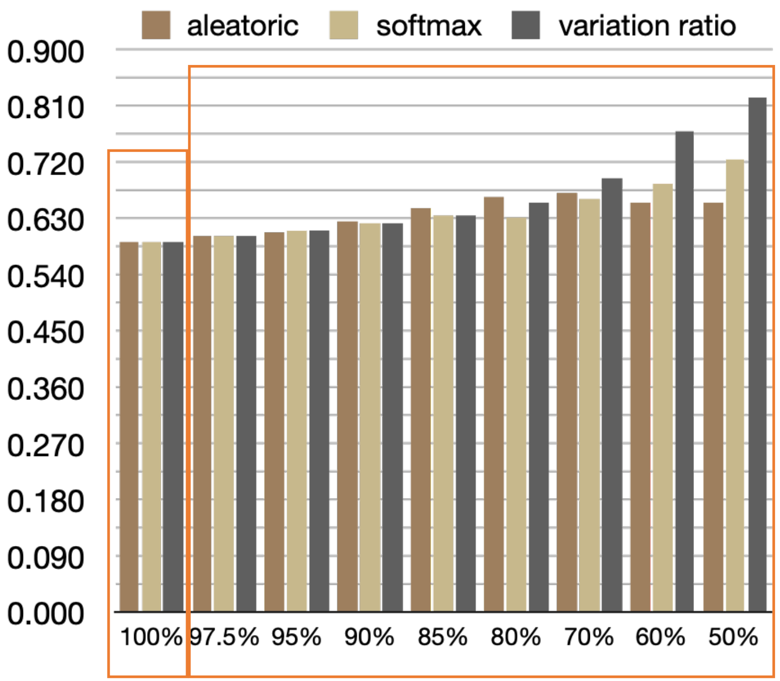
Method from (Kendall and Gal, 2017)

*Kendall, Alex, and Yarin Gal. "What uncertainties do we need in bayesian deep learning for computer vision?." Advances in neural information processing systems. 2017.*

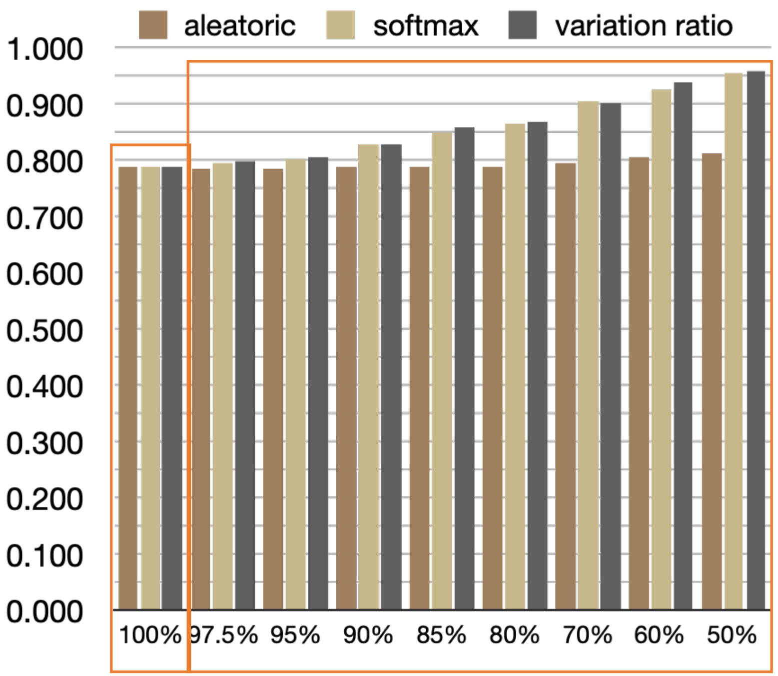
# Unsupervised rejection




PHEME

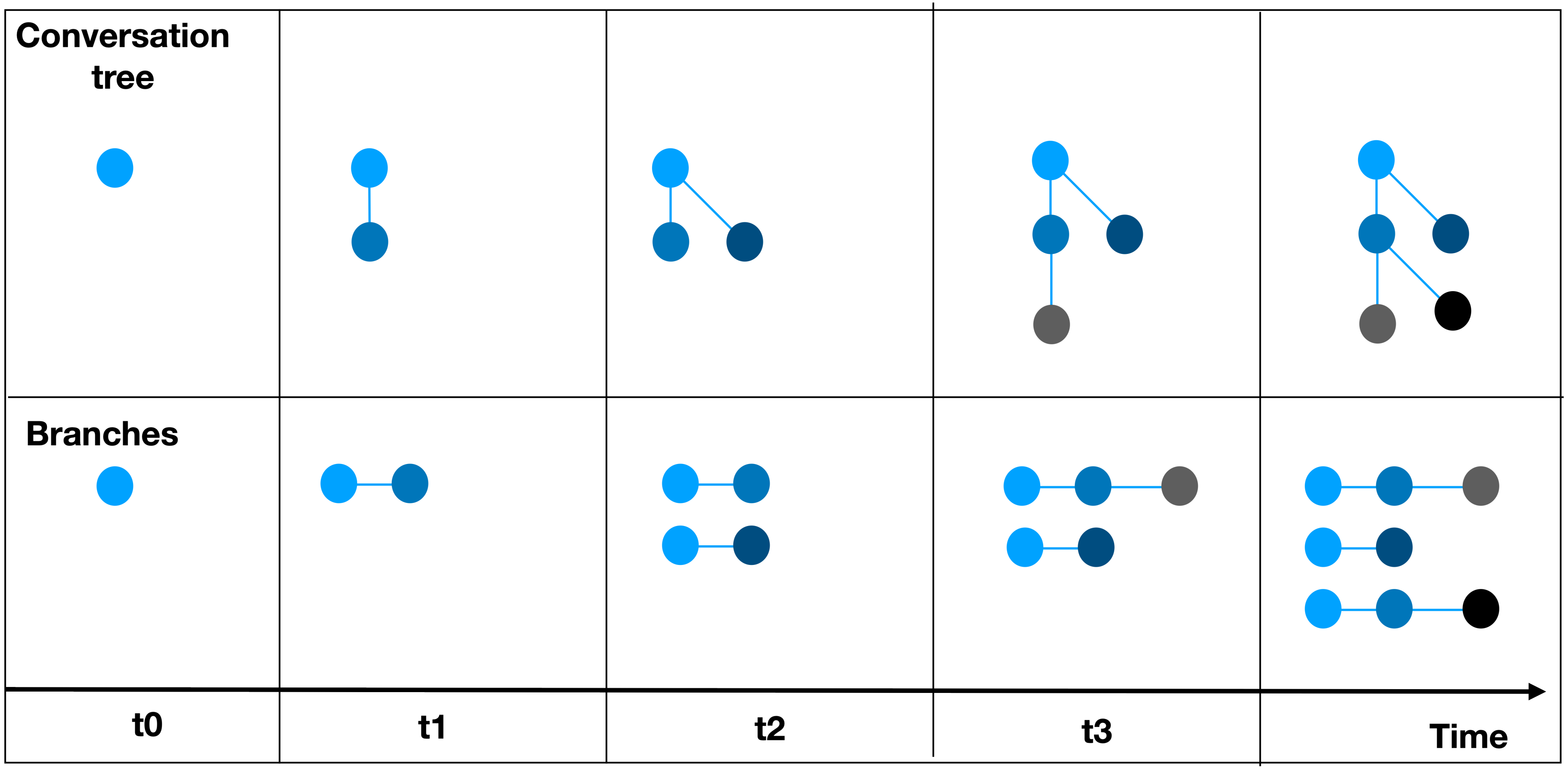



Twitter 15

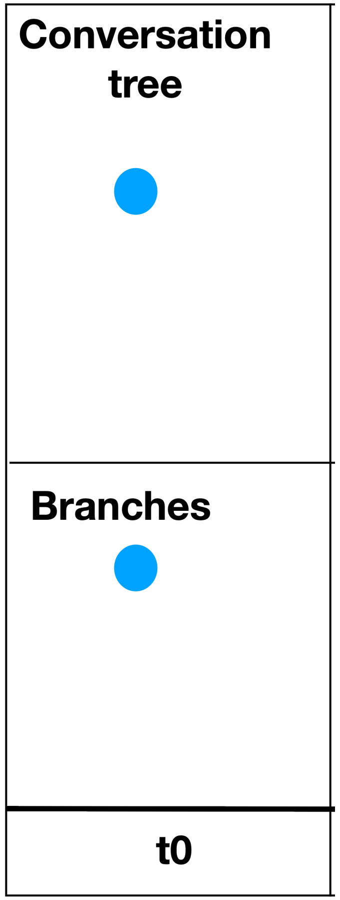



Twitter 16

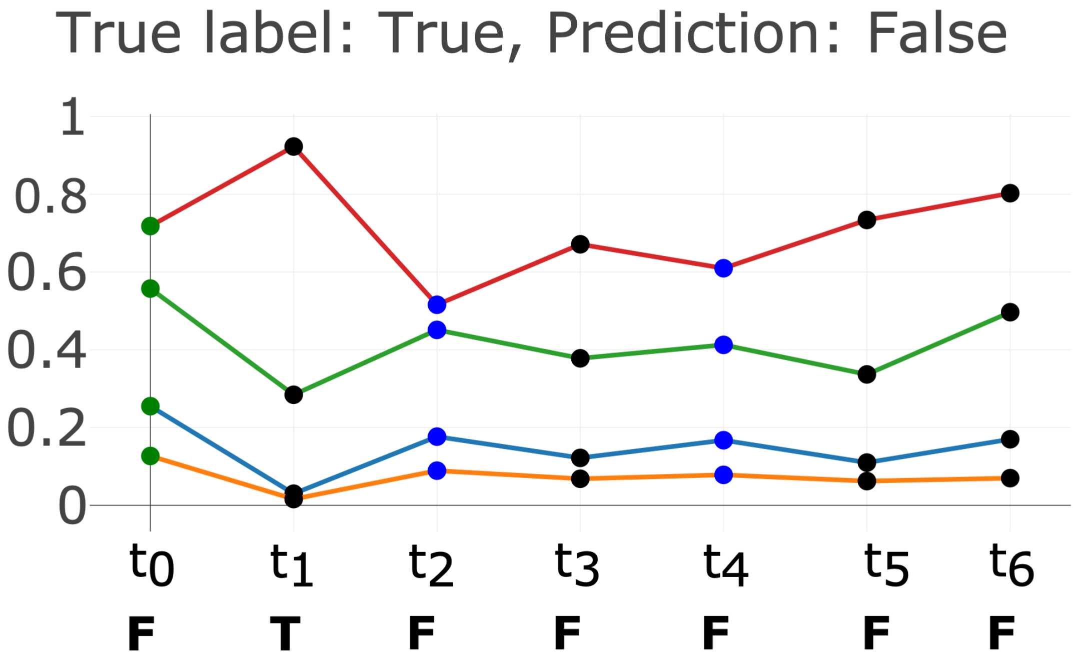
# Timeline experiments



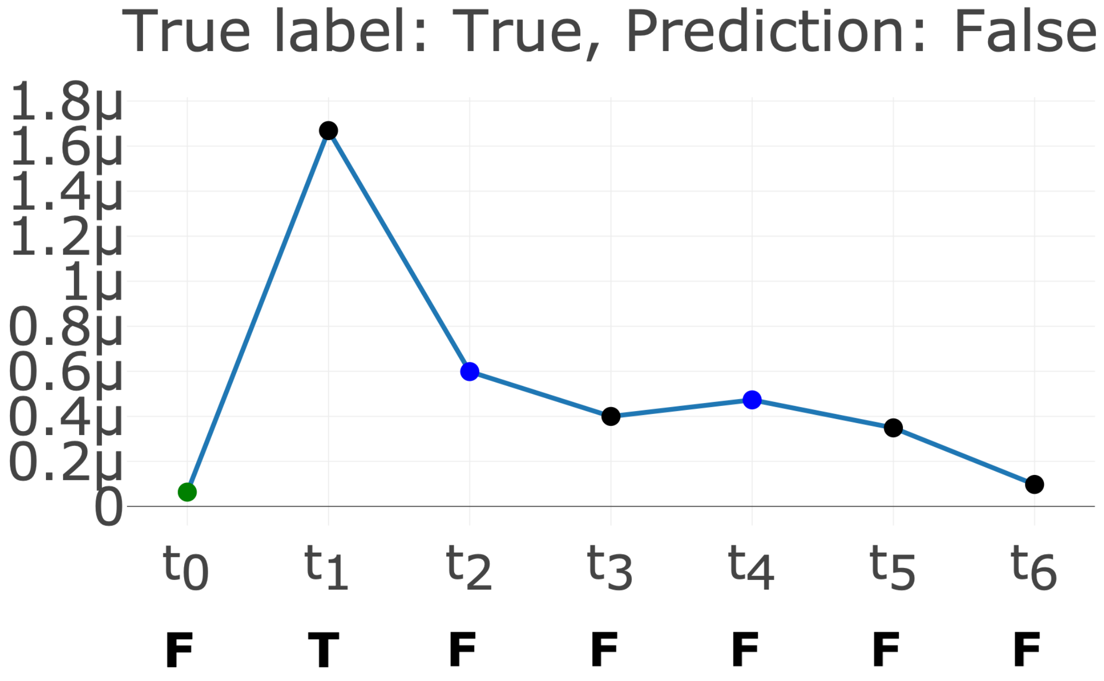
# Timeline experiments



# Timeline experiments




(a) softmax, var.ratio, variance, entropy




(b) aleatoric uncertainty

Nodes: Support - green, Deny - red, Question - blue, Comment - black

# Conclusions

WARWICK

- **Uncertainty estimates can be leveraged to remove instances** that are likely to be incorrectly predicted to **be prioritised by a human fact-checker**.

- Uncertainty estimates can be used to interpret **model decisions over time**.

# Future work

- **Comparison with other methods** for uncertainty estimation
- **Incorporating uncertainty to affect model decisions**, e.g active learning set up
- Further **investigating links between uncertainty** values **and linguistic features** of the input

# Goals and challenges

**The ideal verification model should:**

- Be accurate.
- Generalise to unseen rumours.
- Resolve rumours at an early stage.
- Provide justification and/or explanations for its predictions.
- <u>Not</u> exhibit significant biases.


Email: E.Kochkina@warwick.ac.uk

Twitter: @Elena_Kochkina

Website: https://warwick.ac.uk/ekochkina


THANK YOU!

The Alan Turing Institute

WARWICK

THE UNIVERSITY OF WARWICK