Experiments..

Baselines

- To validate the effectiveness of EANN, choose baselines from the following three cat.:
  - Single Modality Models
    - **Text / Vis**
  - Multi-Modal Models
    - **VQA / NeuralTalk / att-RNN**
  - Variant of the proposed Model
    - **EANN-** (w/o the event discriminator)

Experiments...

Performance Comparison: Twitter Dataset

Method	Accuracy	Precision	Recall	F1
Text	0.532	0.598	0.541	0.568
Vis	0.596	0.695	0.518	0.593
VQA	0.631	0.765	0.509	0.611
NeuralTalk	0.610	0.728	0.504	0.595
att-RNN	0.664	0.749	0.615	0.676
EANN-	0.648	0.810	0.498	0.617
EANN	**0.715**	**0.822**	**0.638**	**0.719**

- # of Tweets on different events is imbalanced and more than 70% of tweets are related to a single event.
  - Cause the learned the text feature mainly focus on some specific events.
  - Seriously prevent extracting transferable feature among events on **Text** Model
- **Text** is lowest, **Vis** is better than **Text**.
  - Images are more transferable, with VGG19 extracting useful feature.
  - **Vis** still worse than that multi-modal approaches
    - Confirms that multiple modalities is superior for the task of fake news detection.