innovate achieve lead

Where is your Evidence: Improving Fact-checking by Justification Modeling

Tariq Alhindi and **Savvas Petridis**
Department of Computer Science
Columbia University

Smaranda Muresan
Department of Computer Science
Data Science Institute Columbia University

Group Members:
Nalin Deepak (2018A7PS0223P)
Kalit Naresh Inani (2018A7PS0207P)
Yash Bansal (2019A7PS0484P)
Harsh Mahajan (2019A7PS0036P)

Research Paper Implemented

- Fake news detection and fact checking require a large amount of human labelled data to get good results from machine learning algorithms.

- In 2017, Wang introduced the LIAR dataset, which includes 12.8K human labeled short statements from POLITIFACT.COM's API5, and each statement is evaluated by a POLITIFACT.COM editor for its truthfulness.

- **T**he LIAR-PLUS dataset extends the LIAR dataset by **automatically extracting the justification sentences provided by humans in the fact-checking articles.**

- How do we get the justification for a particular fact/claim?
  - The end paragraph usually provides us the summary of the PolitiFact article.
  - Phrases containing "**our ruling**" and "**summing up**" aid to get the justification.

Problem Description

- **Basis:** The metadata features, sentiment strength and emotions, extracted justification can support or contradict the claim, and help improving the performance of the model.

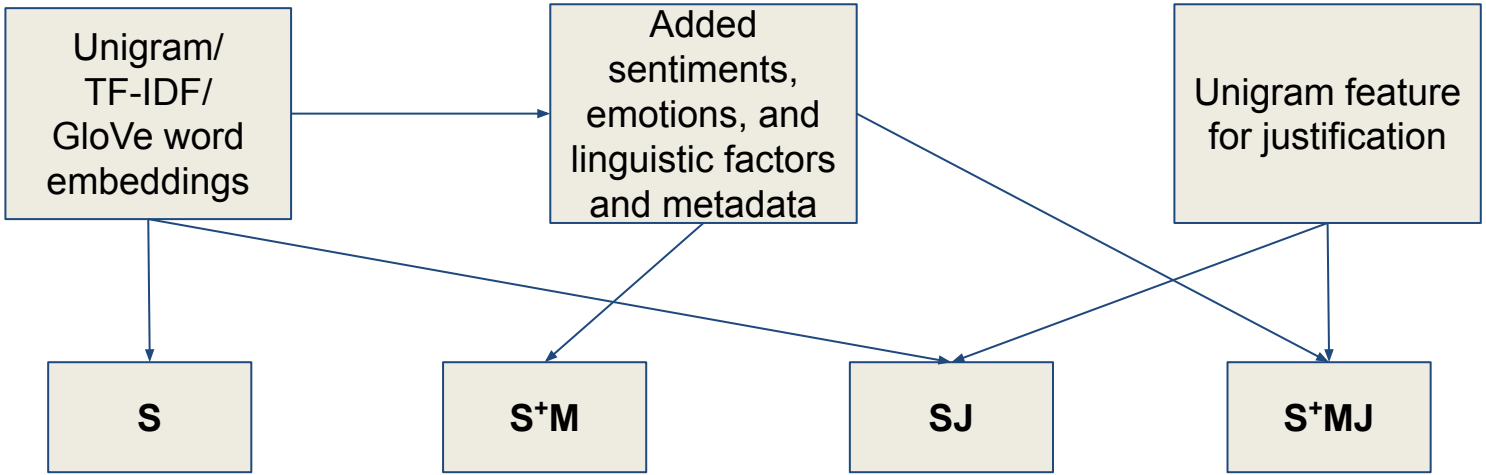
- **Conditions:**
  - **S condition**- basic claim/statement representation using just word representations
  - **S+M condition**- enhanced claim/statement representation that captures additional information shown to be useful sentiment strength and emotion as well as metadata information
  - **SJ condition** basic claim/statement and the associated extracted justification
  - **S+MJ condition**- enhanced claim/statement representation, metadata and justification

- **Classification Categories**:
  - **Binary classification task** classifies into either true or false
  - **Six-way classification task** classifies into “pants on fire”, “false”, “mostly false”, “half-true”, “mostly true” or “true”.

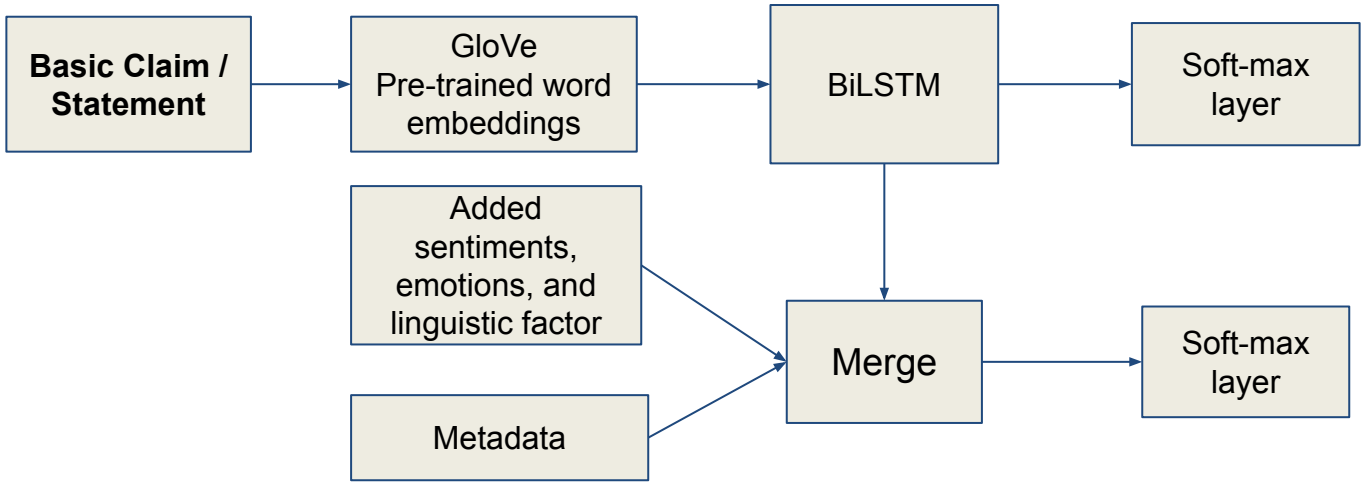
Word representation techniques

Representation	Model	Six way		Binary	
		val	test	val	test
unigram	LR	0.2346	0.2305	0.5737	0.619
	SVM	0.2228	0.2283	0.5684	0.5801
tf-idf	LR	0.2303	0.2554	0.606	0.6266
	SVM	0.2314	0.2294	0.578	0.6212
glove	LR	0.2521	0.2271	0.6028	0.6006
	SVM	0.2532	0.2271	0.605	0.6028
word2vec	LR	0.2258	0.2328	0.6214	0.5974
	SVM	0.2461	0.2517	0.6214	0.6069

Feature Based Algorithms - LR and SVM



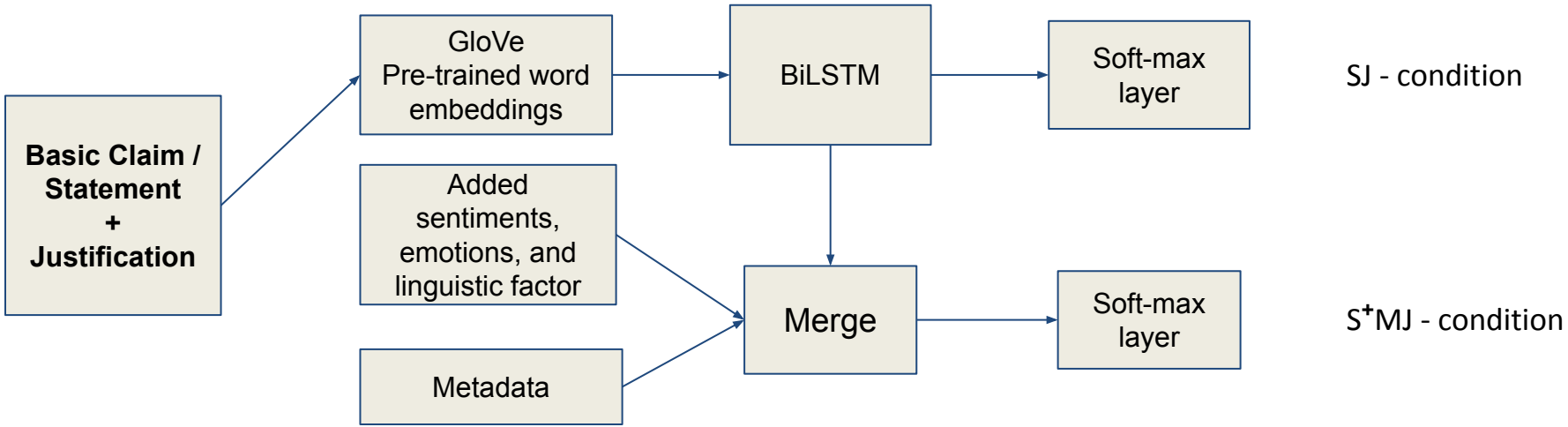
Deep Learning Models



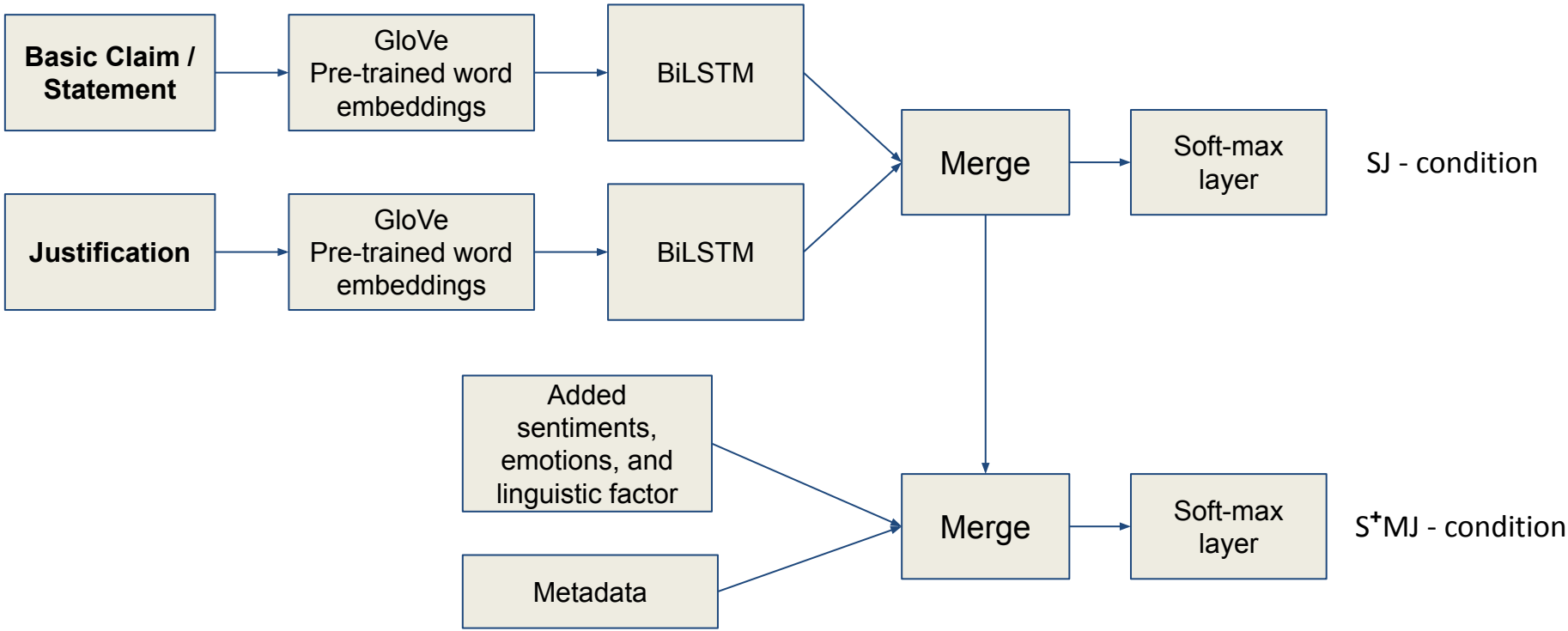
S - condition

$S^{+}M$ - condition

Deep Learning Models



Deep Learning Models - P-BiLSTM



Results solutions proposed in the paper

Cond.	Model	Binary		Six-way	
		valid	test	valid	test
S	LR	0.58	0.61	0.23	0.25
	SVM	0.56	0.59	0.25	0.23
	BiLSTM	0.59	0.60	0.26	0.23
SJ	LR	0.68	0.67	0.37	0.37
	SVM	0.65	0.66	0.34	0.34
	BiLSTM	0.70	0.68	0.34	0.31
	P-BiLSTM	0.69	0.67	0.36	0.35
$S^{+}M$	LR	0.61	0.61	0.26	0.25
	SVM	0.57	0.60	0.26	0.25
	BiLSTM	0.62	0.62	0.27	0.25
$S^{+}MJ$	LR	0.69	0.67	0.38	0.37
	SVM	0.66	0.66	0.35	0.35
	BiLSTM	0.71	0.68	0.34	0.32
	P-BiLSTM	0.70	0.70	0.37	0.36

Cond	Model	Binary		Six-way	
		valid	test	valid	test
S	LR	0.5737	0.619	0.2346	0.2305
	SVM	0.5684	0.5801	0.2228	0.2283
	BiLSTM	0.61	0.622	0.244	0.247
SJ	LR	0.564	0.5823	0.2368	0.237
	SVM	0.5576	0.5758	0.2336	0.2067
	BiLSTM	0.561	0.567	0.248	0.24
	P-BiLSTM	0.576	0.582	0.252	0.253
S+M	LR	0.6329	0.6212	0.2368	0.237
	SVM	0.5899	0.6071	0.2336	0.2067
	BiLSTM	0.554	0.54	0.267	0.264
S+MJ	LR	0.6136	0.6126	0.2626	0.2273
	SVM	0.5856	0.5942	0.2443	0.2089
	P-BiLSTM	0.61	0.622	0.255	0.247

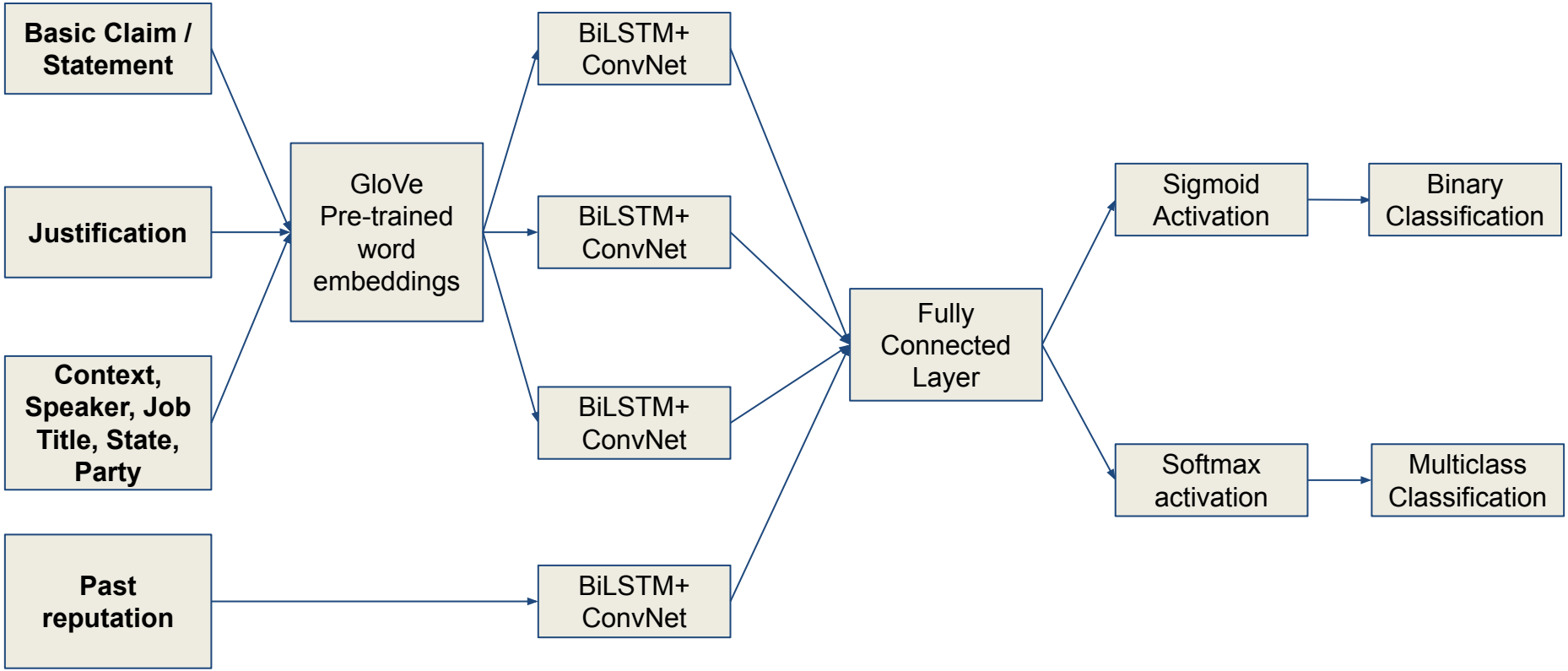
Our Solutions

An issue we faced is to find a robust method to model justification and metadata along with the statement. We tried out various approaches for this, but a significant increase in performance was not observed when using the architectures and methods specified in the paper. This may be due to:

a. Nonoptimal way of concatenating/combining justification, metadata, and other additional information along with the statement.
b. Bad tuning of hyperparameters of the model.

Hence we implemented the following two approaches as well which gave us better accuracies for the task conditions but not necessarily the desired results

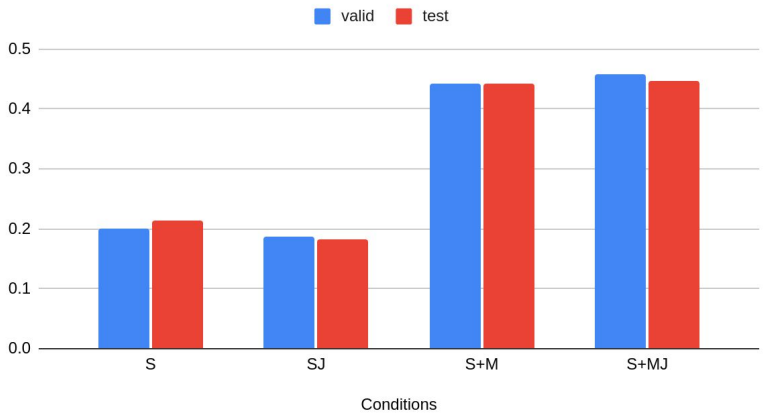
Improved LSTM Technique



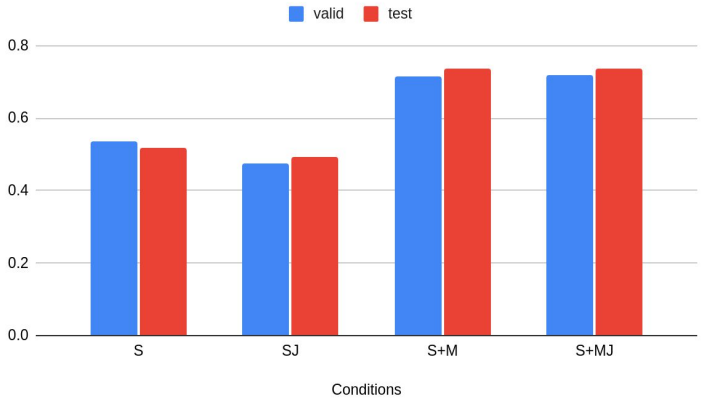
Our results

Conditions	Six-way		Binary	
	valid	test	valid	test
S	0.200935	0.213891	0.537383	0.516969
SJ	0.186137	0.18311	0.475857	0.492502
S+M	0.442368	0.442778	0.715732	0.734807
S+MJ	0.457165	0.447514	0.71729	0.736385

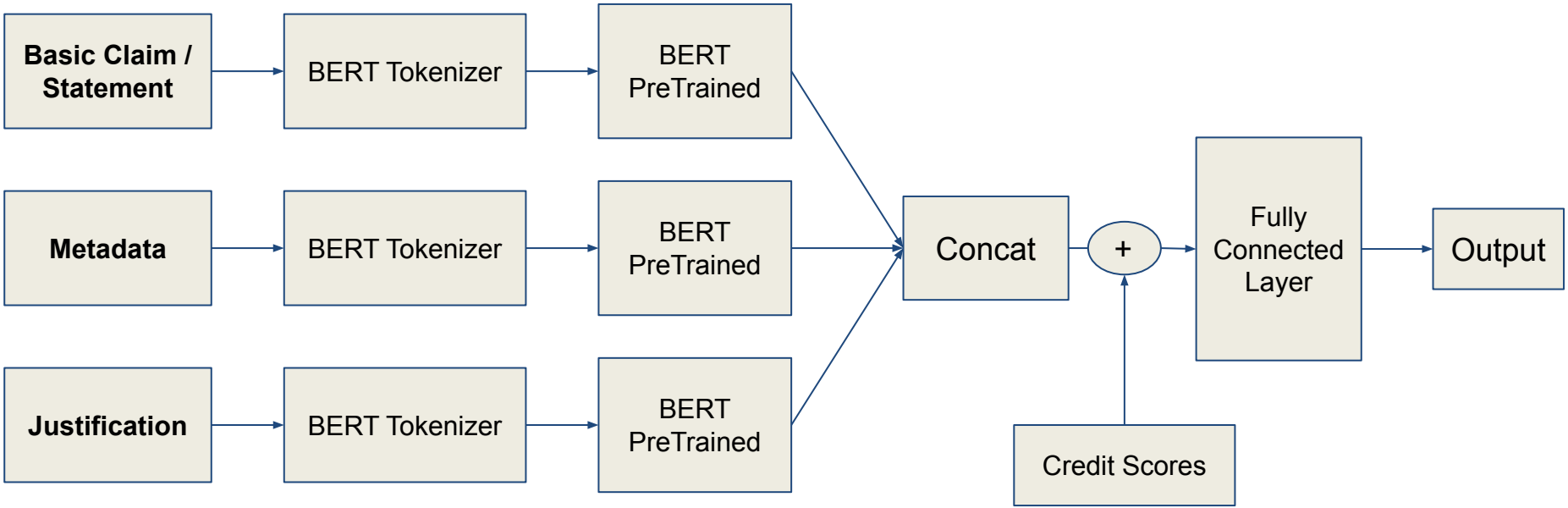
Multi class classification



Binary classification

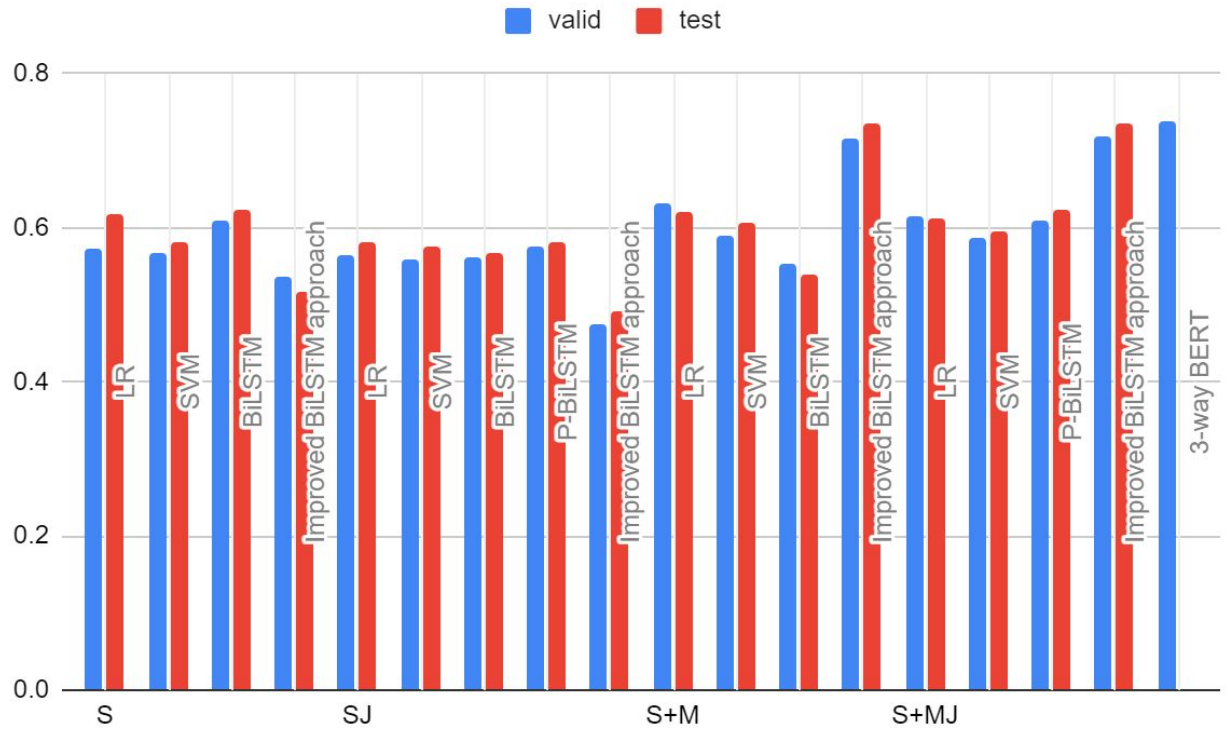


Feature extraction from BERT

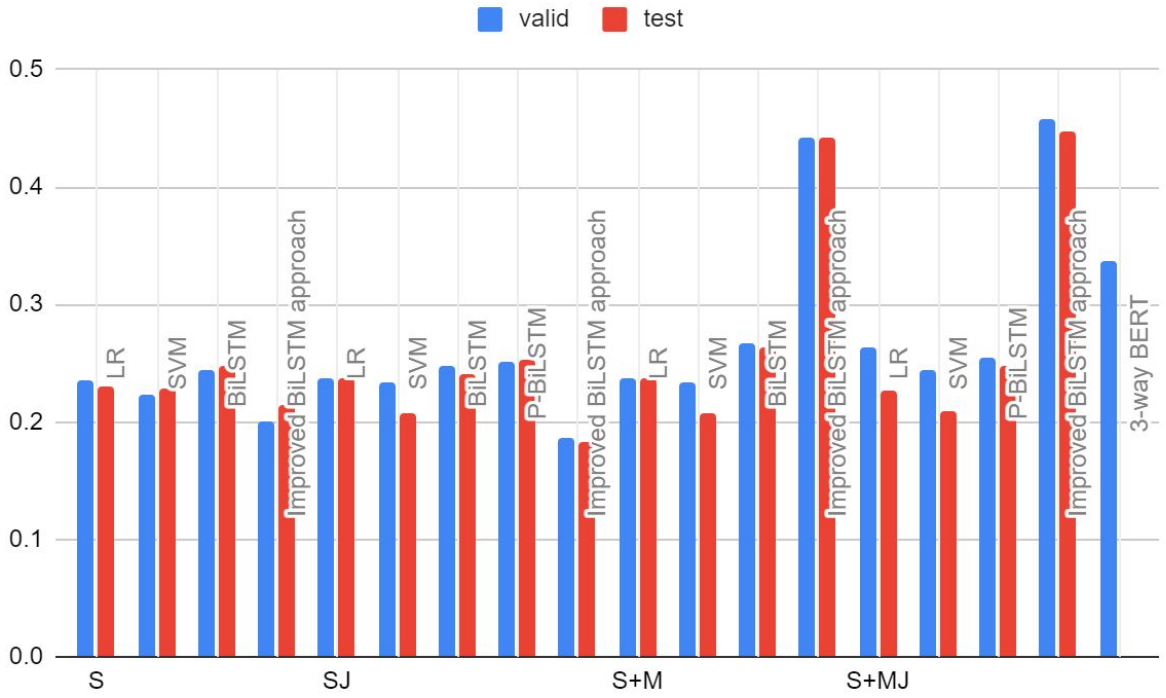


Binary Classification : 0.7395
Six-Way Classification : 0.3370

Overall Results Comparisons - Binary



Overall Results Comparisons - Six-Way



Conclusion

- We tried to implement various approaches given in the paper, and verify the effect of adding justification to the inputs.
- Unfortunately we could not reproduce the significant rise in accuracy with justification using the given approaches, hence we tried the improved Bi-LSTM + ConvNet based architecture as well as a 3 way BERT approach.
  - These gave us better results in terms of accuracy but not necessarily the results we were expecting
- **Future scope**
  - Refinement of justification extraction method
  - Methods for evidence extraction from the web
  - Robust techniques to concatenate metadata and justification with statement