

Natural Language Inference

Team 25

Members

- Abhay Kaushik 2022201054
- Rohit Hooda 2022201034
- Santanu Biswas 2022201031



Datasets

01

SNLI dataset
(Stanford Natural
Language
Inference)

02

MultiNLI dataset
(Multi-Genre
Natural
Language
Inference)

03

ANLI dataset
(Adversarial
Natural
Language
Inference)



Comparison Of Datasets

01

SNLI

Train Set – 550k

Test Set – 10k

Dev Set – 10k

02

MultiNLI

Train Set – 392k

Test Set – 20k

Dev Set – 20k

03

ANLI

Train Set – 45.5k

Test Set – 1k

Dev Set – 1k

Models

MLP / Sum of Words Model

LSTM and BiLSTM Models

Decomposable Attention Model

ELMo Model

BERT Model

Pre-processing



Tokenization and
cleaning

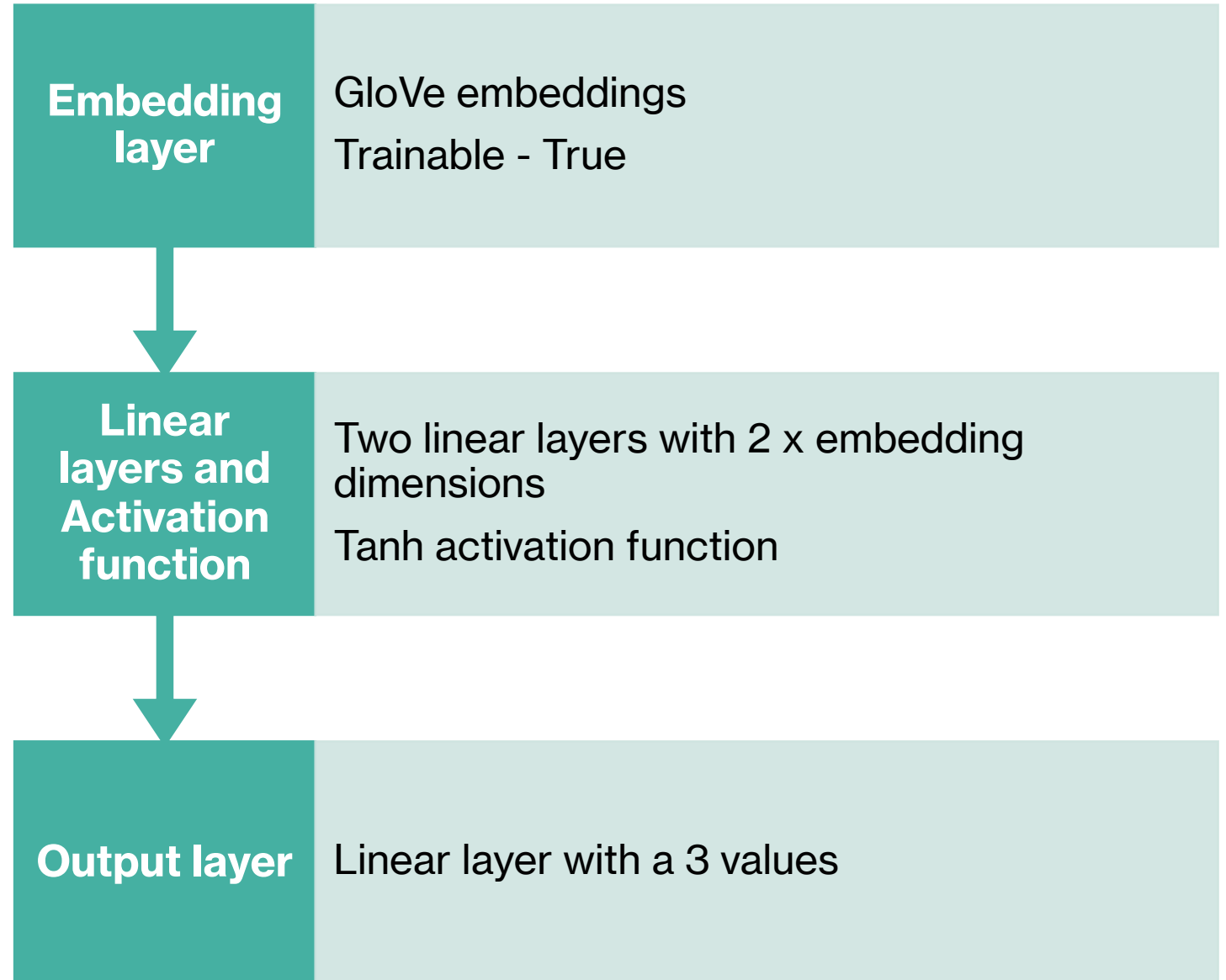


Word2index
mapping



Creating datasets
and dataloader

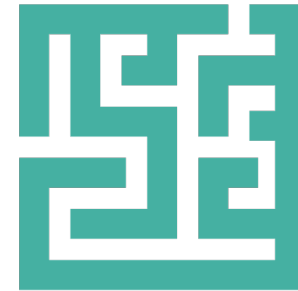
Sum of Words/MLP Model



Observations / Results / Analysis

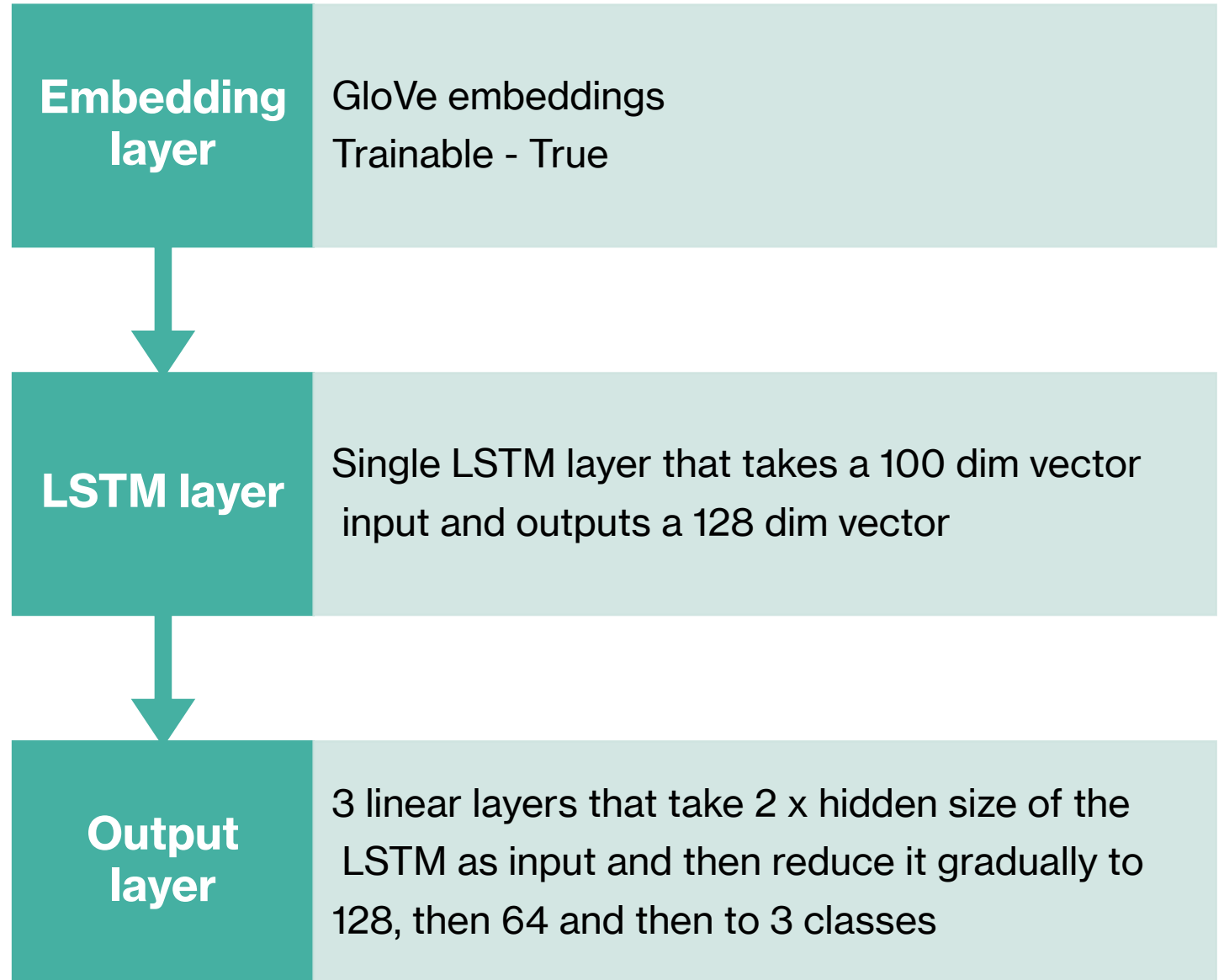


Unable to capture information complexity.

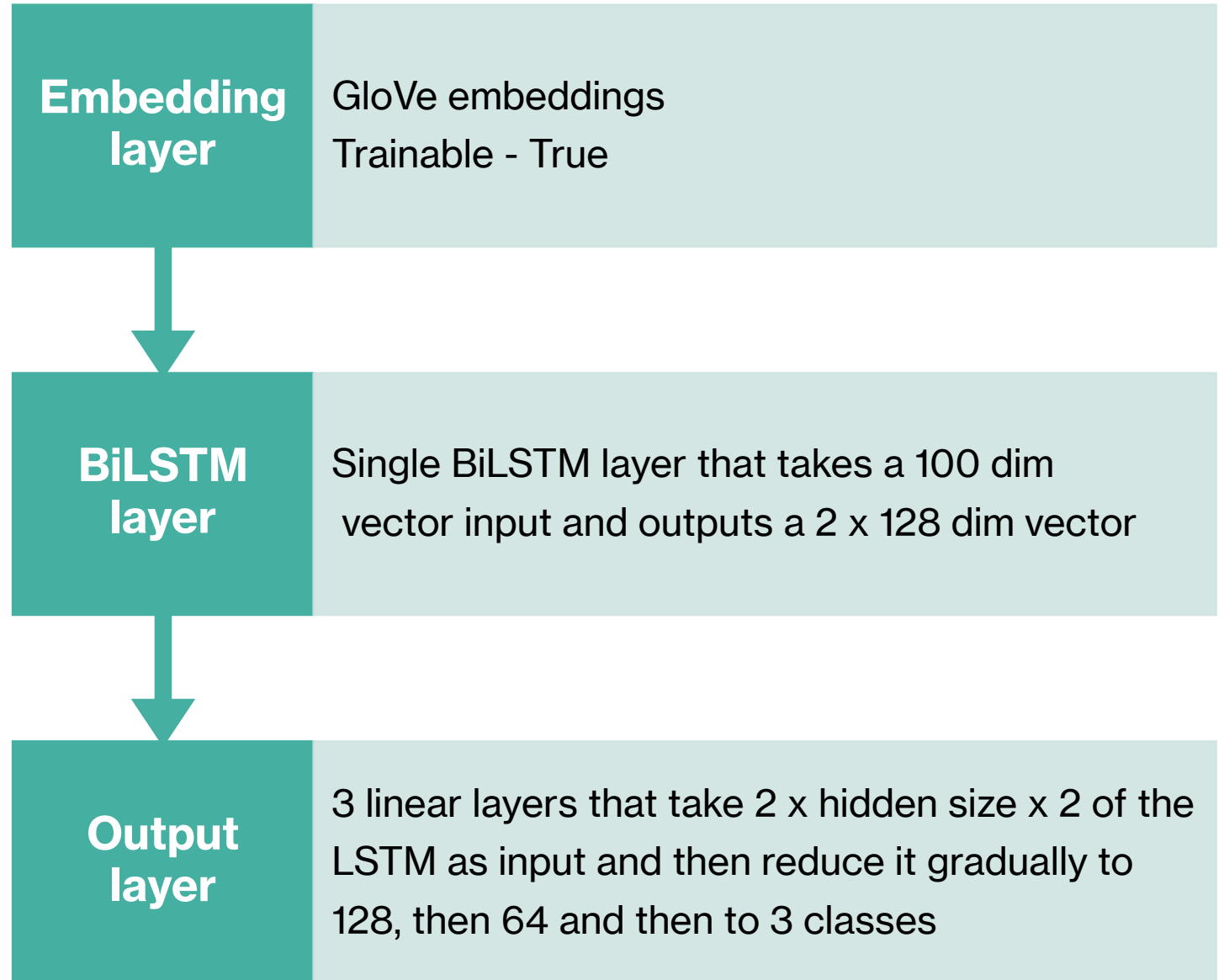


More complex model might be needed ?

LSTM Model



BiLSTM Model

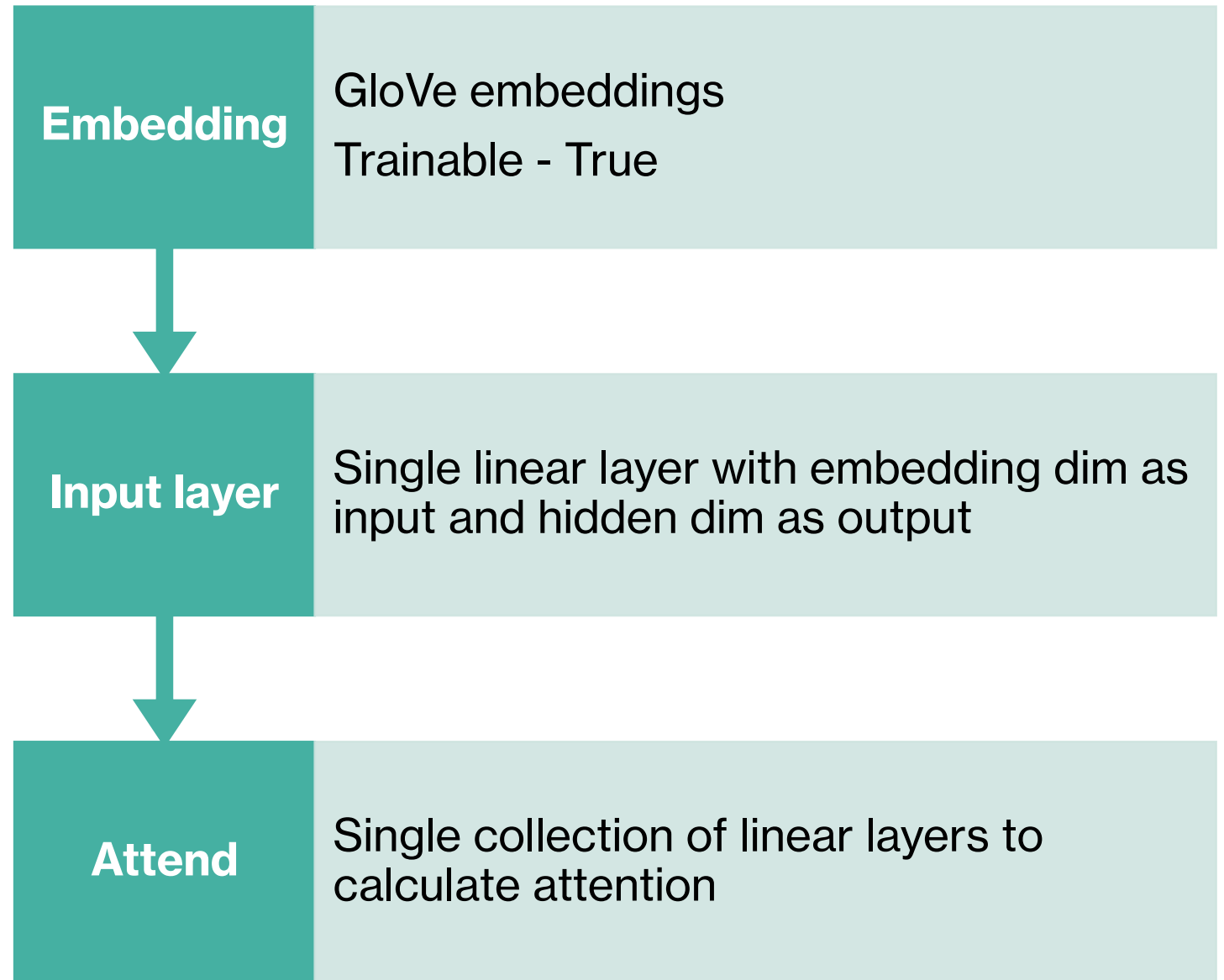




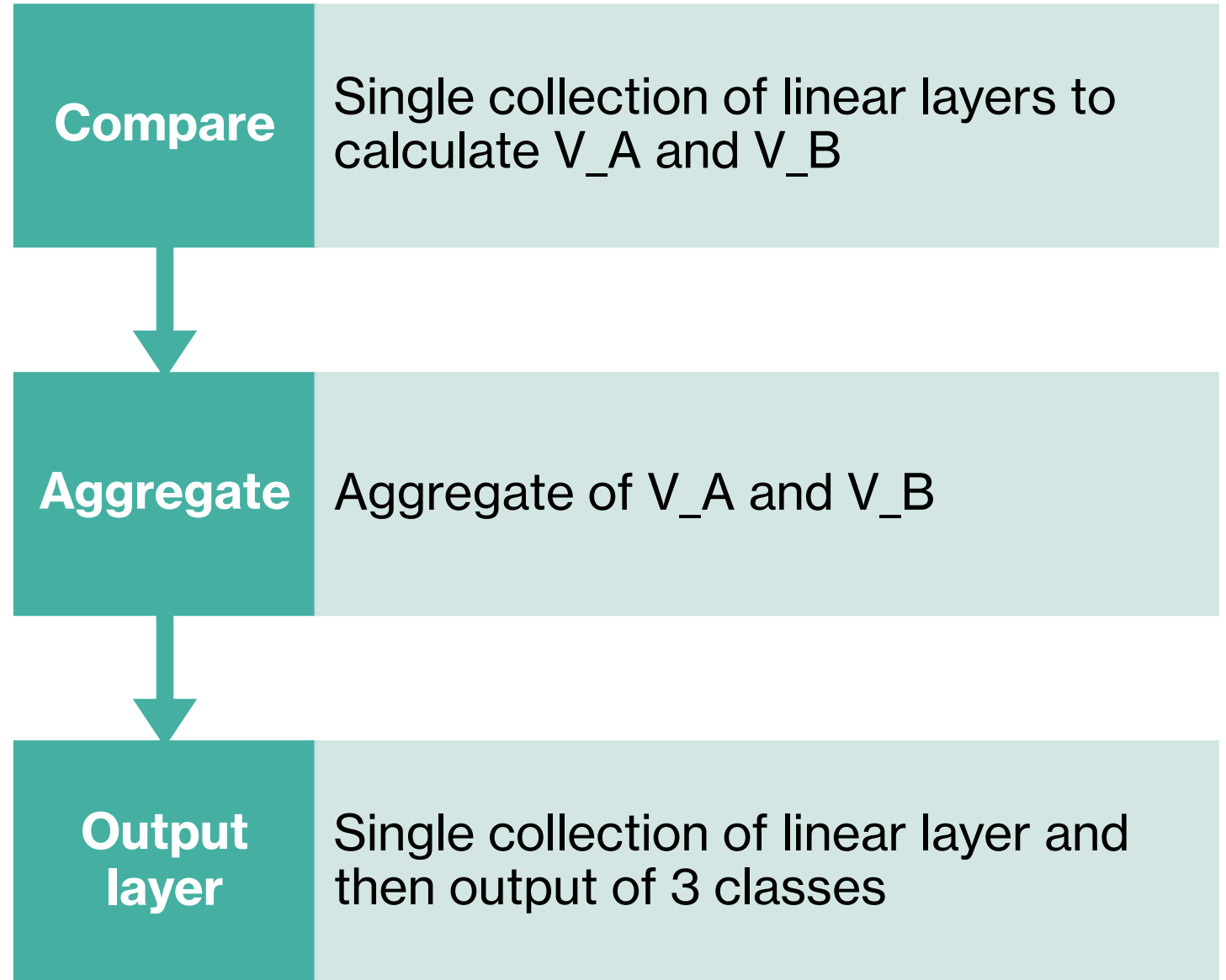
Observations / Results / Analysis

- Faster convergence due to Glove embeddings.
- Better than MLP at capturing patterns.
- Could be improved if a larger corpus is used.
- BiLSTM Model gives better loss than LSTM.

Attention Model



Attention Model

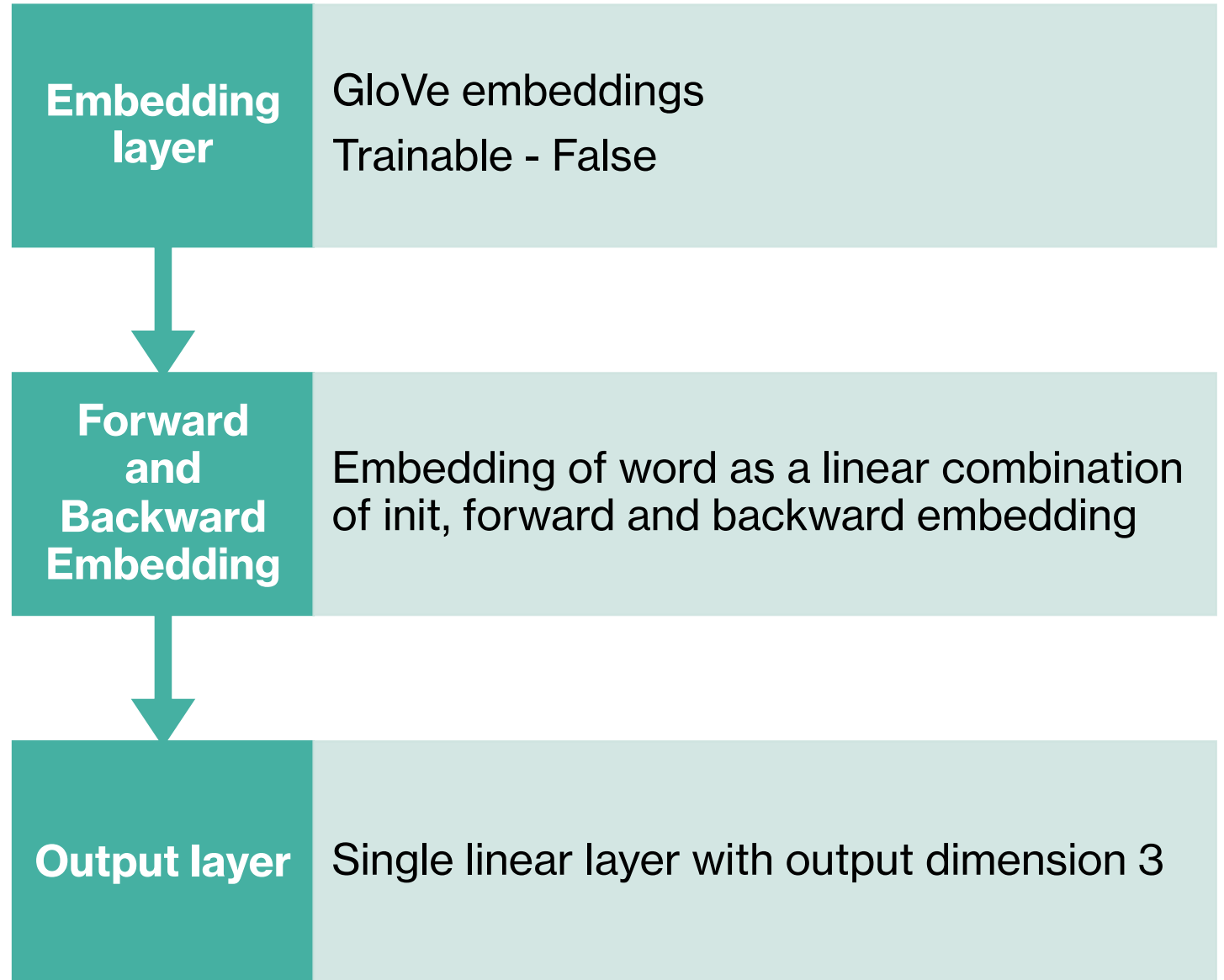


Observations / Results / Analysis

- Performance weaker than that of LSTM but better than MLP.
- One reason is that LSTM is capturing the sequential nature.



ELMo Model



Observations / Results / Analysis

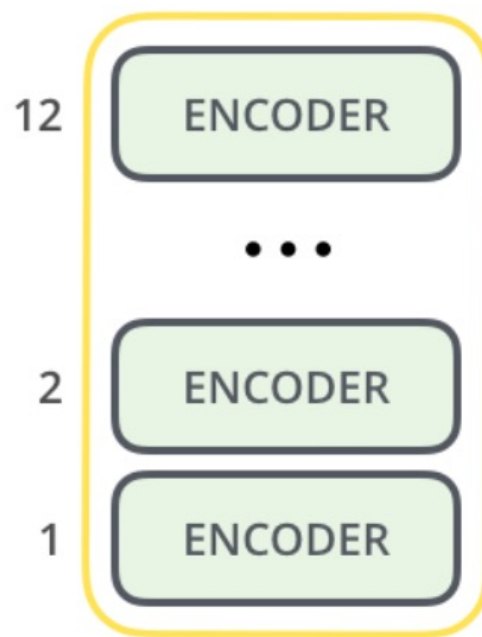
- Not as good as EMLo could be –
 - Reason – Not trained at character level
- Can't handle out of vocabulary words, while the original ELMo model can.



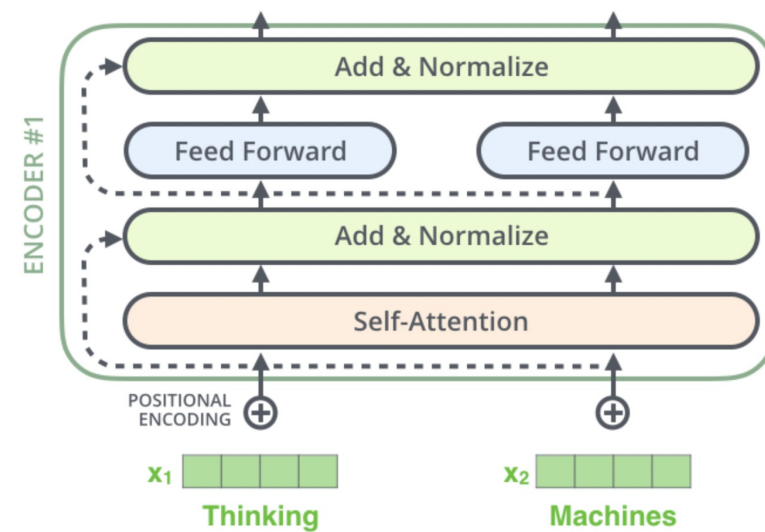
BERT



BERT_{BASE}



BERT_{BASE}





Sample Inference Outputs - SNLI

Premise	Hypothesis	Output	Actual
This church choir sings to the masses as they sing joyous songs from the book at a church	The church is filled with song	entailment	entailment
A woman with a green headscarf, blue shirt and a very big grin.	The woman has been shot.	entailment	contradiction
A man playing an electric guitar on stage	A man playing banjo on the floor	contradiction	contradiction



Sample Inference Outputs - MultiNLI

Premise	Hypothesis	Output	Actual
The new rights are nice enough	Everyone really likes the newest benefits	neutral	neutral
This site includes a list of all award winners and a searchable database of Government Executive articles	The Government Executive articles housed on the website are not able to be searched	contradiction	contradiction
you want to punch the button and go	You don't want to push the button lightly, but rather punch it hard	neutral	neutral

Sample Inference Outputs - ANLI

Premise	Hypothesis	Output	Actual
It is Sunday today, let's take a look at the most popular posts of the last couple of days. Most of the articles this week deal with the iPhone, its future version called the iPhone 8 or iPhone Edition, and new builds of iOS and macOS. There are also some posts that deal with the iPhone rival called the Galaxy S8 and some other interesting stories. The list of the most interesting articles is available below. Stay tuned for more rumors and don't forget to follow us on Twitter	The day of the passage is usually when Christians praise the lord together	contradiction	entailment
The Real Estate Pros are focused on providing you with the best results and service in the industry. They listen carefully to understand your real estate goals and work hard to create solutions that make sense for you. They have extensive knowledge of the Rock Springs, WY area and can help you find the right home for you or the right buyer for your home. Related	There are better real estate companies than Real Estate Pros in Wyoming	contradiction	neutral
Some of the emails had the subject line, "Think Twice." They were sent from a spoofed email address. The sender claimed to have had an associate plant a small bomb in the recipient's building and said the only way to stop him from setting it off was by making an online payment of \$20,000 in bitcoin	The sender would have accepted \$20,000 in any crypto-currency	contradiction	contradiction



BEST OUT OF ALL!

Observations / Results Pt. 1 SNLI

	precision	recall	f1-score	support
entailment	0.89	0.92	0.91	3368
neutral	0.93	0.92	0.93	3237
contradiction	0.87	0.86	0.87	3219
accuracy			0.90	9824
macro avg	0.90	0.90	0.90	9824
weighted avg	0.90	0.90	0.90	9824

Observations / Results Pt. 2 - MultiNLI

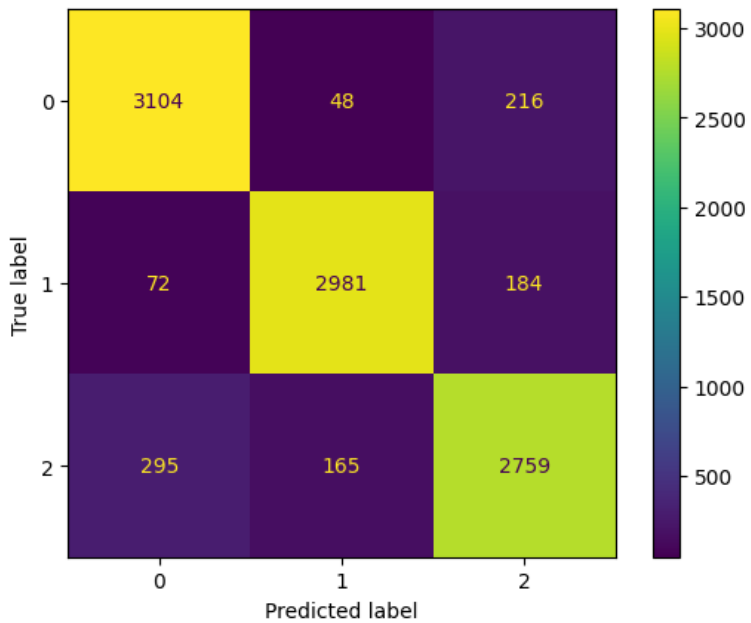
	precision	recall	f1-score	support
entailment	0.86	0.86	0.86	3479
neutral	0.85	0.86	0.85	3213
contradiction	0.79	0.79	0.79	3123
accuracy			0.84	9815
macro avg	0.83	0.84	0.84	9815
weighted avg	0.83	0.84	0.84	9815

Observations / Results Pt. 3 – ANLI

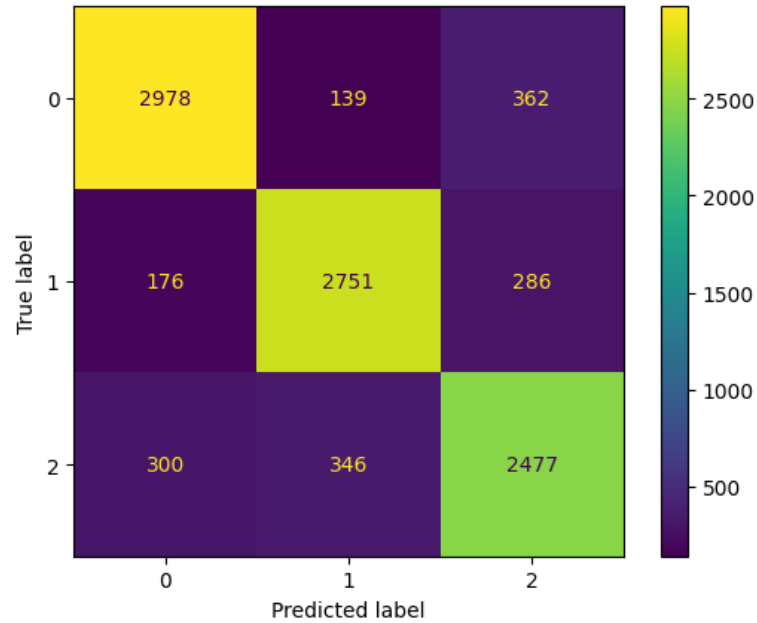
	precision	recall	f1-score	support
entailment	0.73	0.74	0.73	4464
neutral	0.65	0.59	0.62	3111
contradiction	0.52	0.56	0.54	2425
accuracy			0.65	10000
macro avg	0.63	0.63	0.63	10000
weighted avg	0.63	0.65	0.65	10000

Confusion Matrices

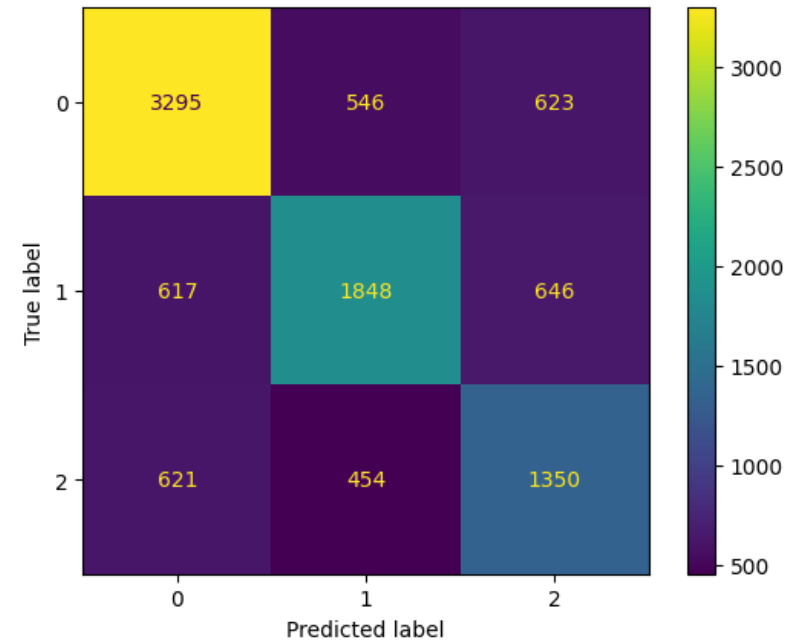
SNLI



MultiNLI



ANLI





Analysis

- Best among all the models tried.
- Still quite low score for ANLI-
 - Reason - Dataset curated so as to create a new benchmark.

Conclusion

Accuracy Score of different models

	MLP	LSTM	BiLSTM	Attn Model	ELMo	BERT
--	-----	------	--------	---------------	------	------

SNLI	73	77	76	61	61	90
------	----	----	----	----	----	----

MultiNLI	52 / 52	53 / 54	53 / 54	51 / 51	43 / 43	84
----------	---------	---------	---------	---------	---------	----

ANLI A3	35	36	35	34	33	65
---------	----	----	----	----	----	----

REFERENCES:

- [1] Samuel R. Bowman, Gabor et. al.: A large annotated corpus for learning natural language inference (2015)
- [2] Ankur P. Parikh et. al.: A Decomposable Attention Model for Natural Language Inference (2016)
- [3] Devlin et. al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2018)
- [4] Williams et. al.: A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference (2017)
- [5] Nie et al.: Adversarial NLI: A New Benchmark for Natural Language Understanding (2020)
- [6] Peters et al.: Deep contextualized word representations (2018)





THANK YOU!

