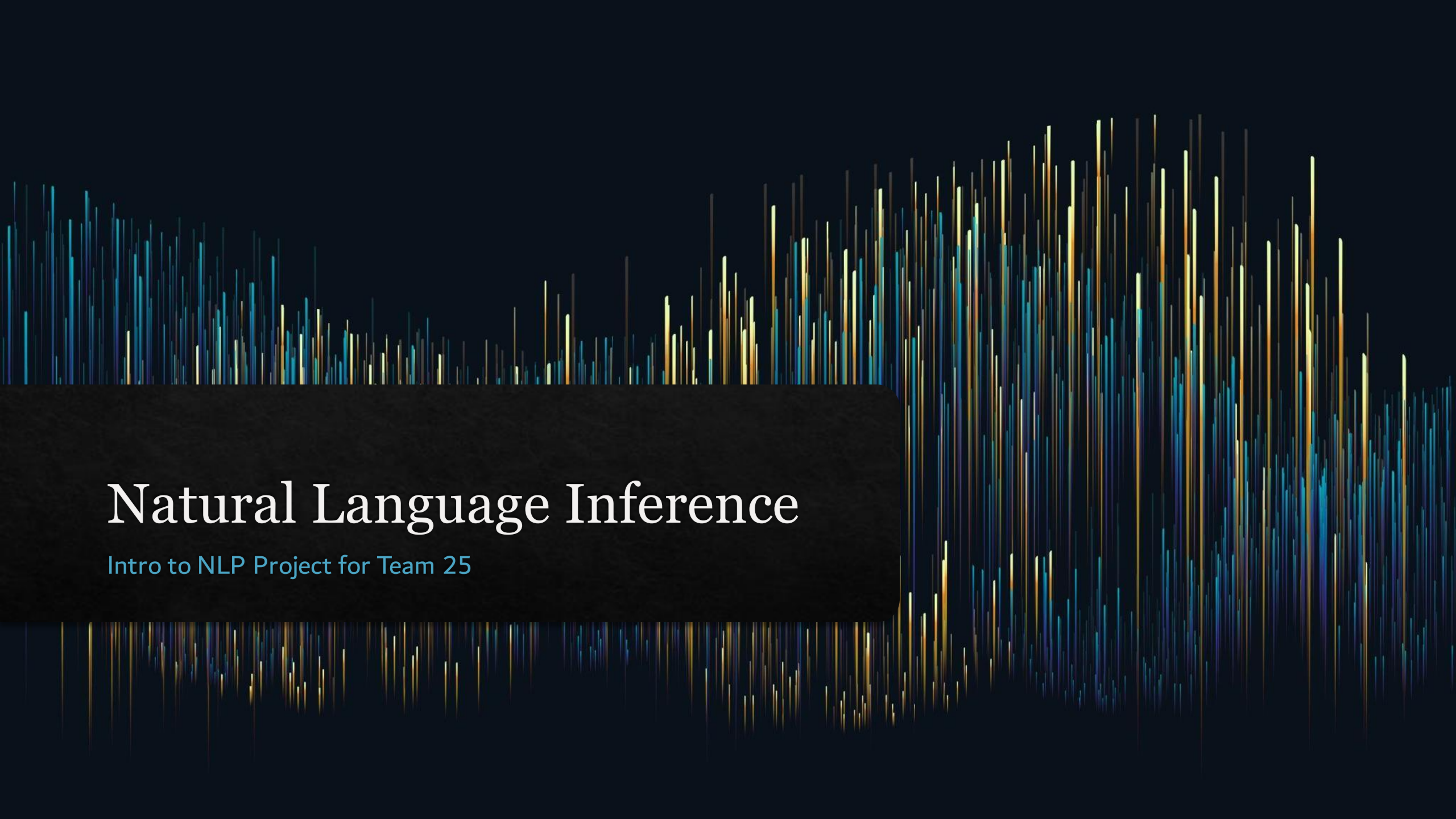
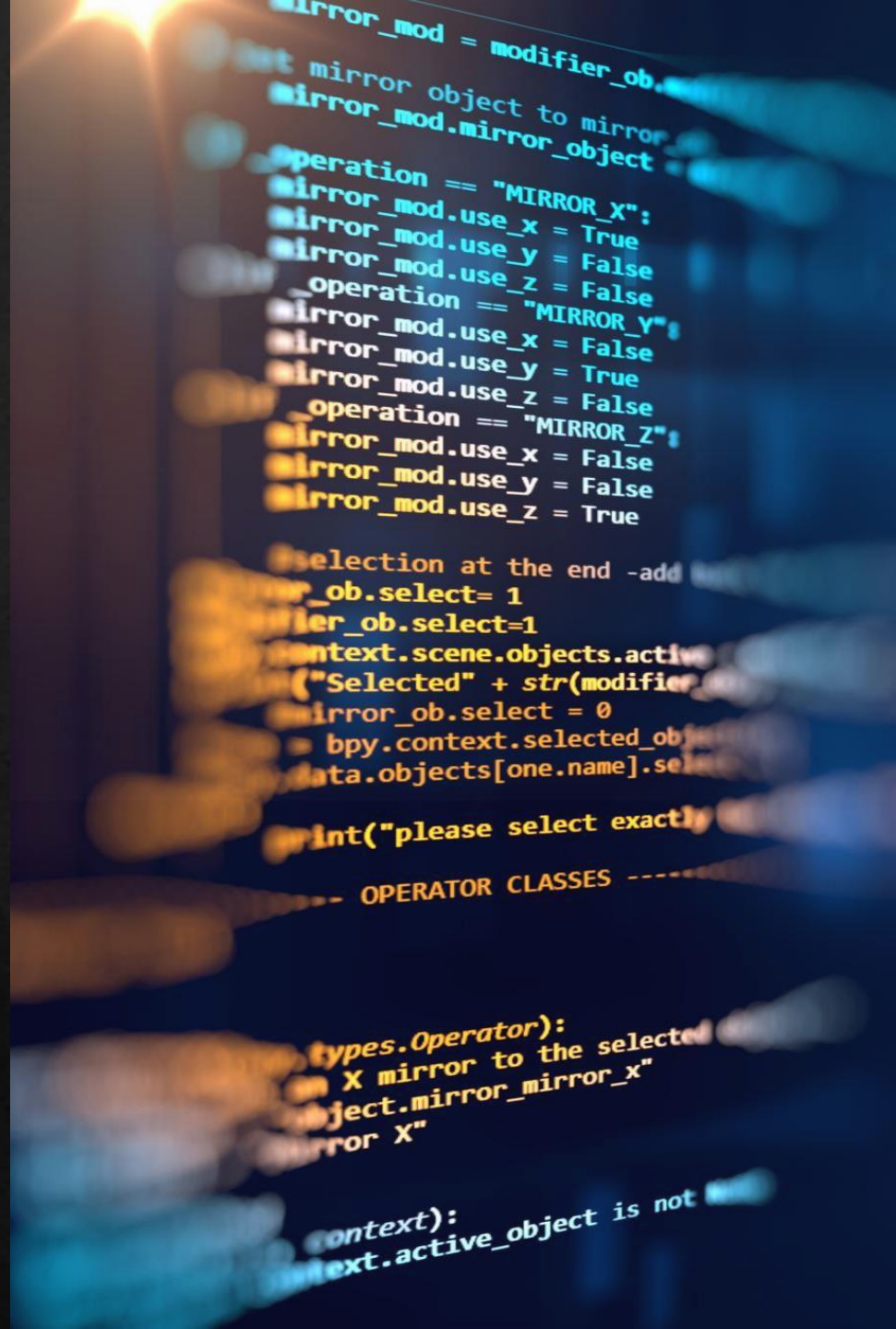


Natural Language Inference

Intro to NLP Project for Team 25

Problem Statement

- Textual entailment recognition decides whether the meaning of one text is entailed from another. TE has a three classifications over sentence pairs:
  - Entailment: texts support each other
  - Neutral: texts unrelated to each other
  - Contradiction: texts oppose each other
- We are to implement a machine learning model for this task.



Relevant Topics

- Natural Language Inference
- Word Embeddings (GloVe, Word2Vec, BERT)
- Attention Model
- RNN and LSTM Mechanism
- Dataset Analysis

Datasets

SNLI: 530k sentence pairs

Human-written sentence pairs manually labelled, by Stanford.

MultiNLI: 433k sentence pairs

Crowd-sourced sentence pairs with more cross-genre generalization than SNLI.

SciTail: 27k sentence pairs

Created from MCQ science exams, where each question answer pair is labelled as entailment.

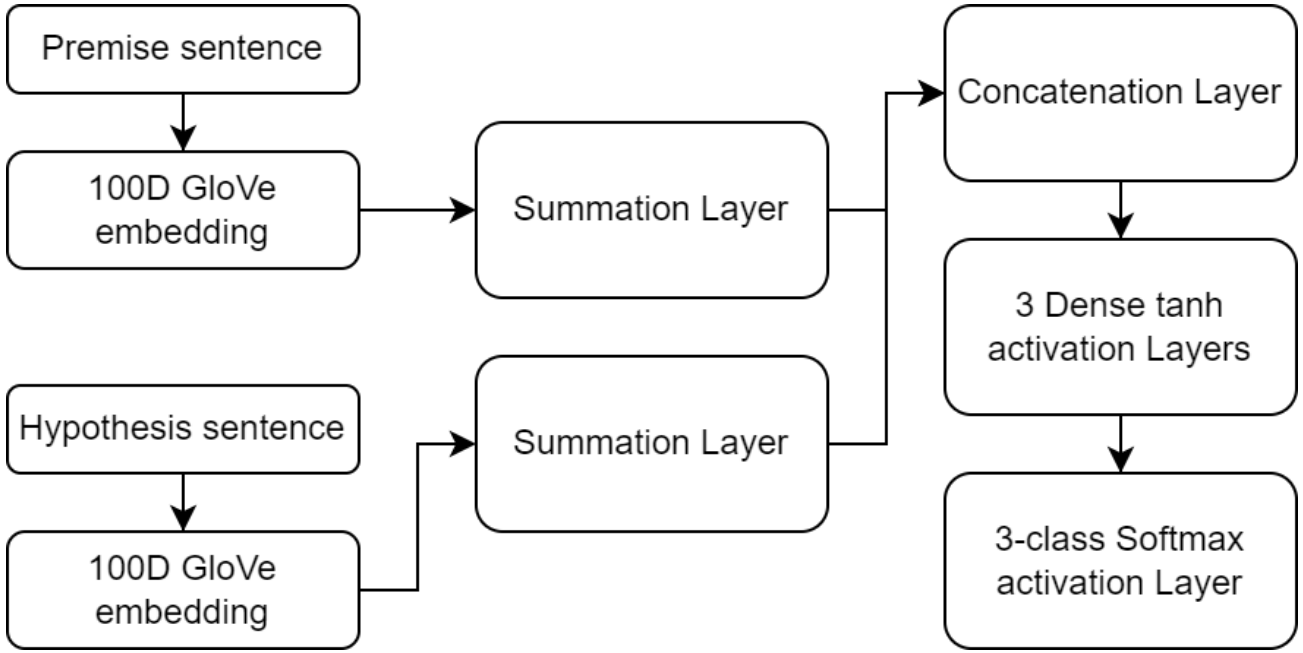
SICK: 10k sentence pairs

Datasets of sentences describing photographs were used as a basis for SICK.

Implementation Specifics

Implementation 1: Bowman Model

- 300D 840B GloVe used to create embedding layer.
- Summation layer: maxLen number of tokens summed to give a 100D representation of both premise and hypothesis.
- Concatenation layer: appends hypothesis to premise creating a 200D vector.
- 200D vector passed through fully connected layers, tanh activation, and dropouts.
- Final fully connected layer: softmax activation over 3 labels.

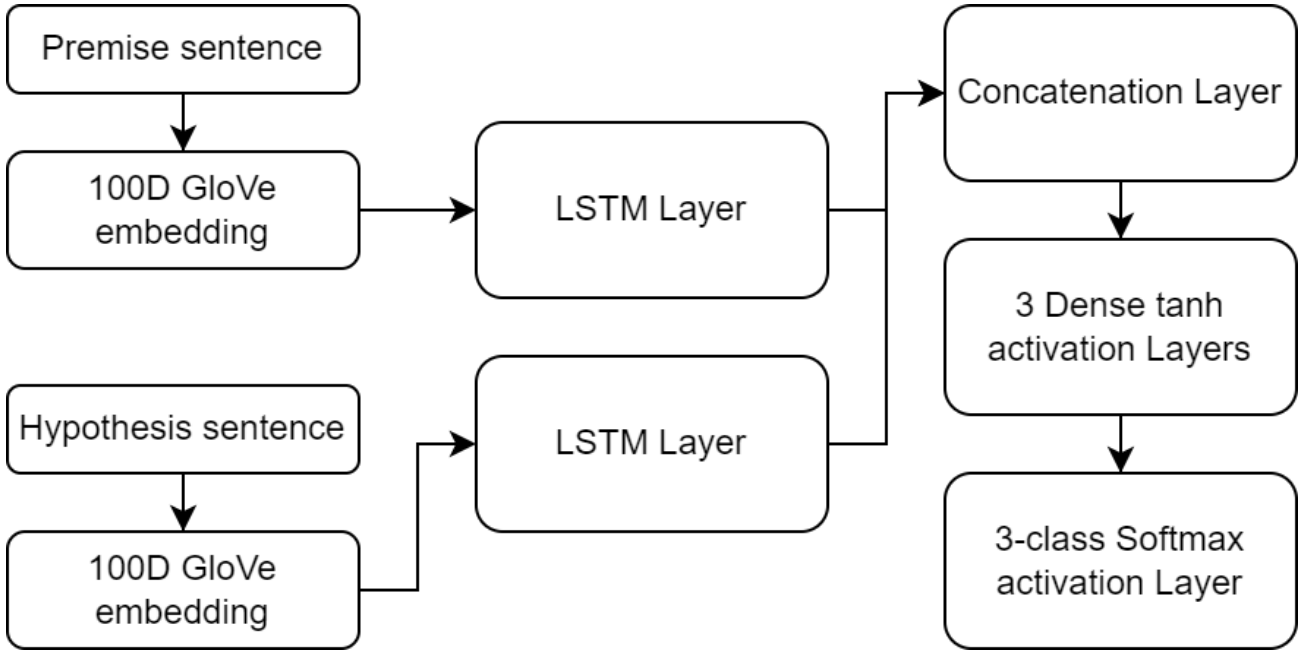


Results

	Training	Validation	Test
SNLI	74%	77%	76%
MultiNLI	65.39%	67.23%	65.74%
SciTail	87.88%	70.68%	71.62%
SICK	88.21%	70.84%	70.49%

Implementation 2: LSTM Model

- 300D 840B GloVe used to create embedding layer.
- LSTM layer: both premise and hypothesis sent to LSTM with an output of 100D vectors.
- Concatenation layer: appends hypothesis to premise creating a 200D vector.
- 200D vector passed through fully connected layers, tanh activation, and dropouts.
- Final fully connected layer: softmax activation over 3 labels.

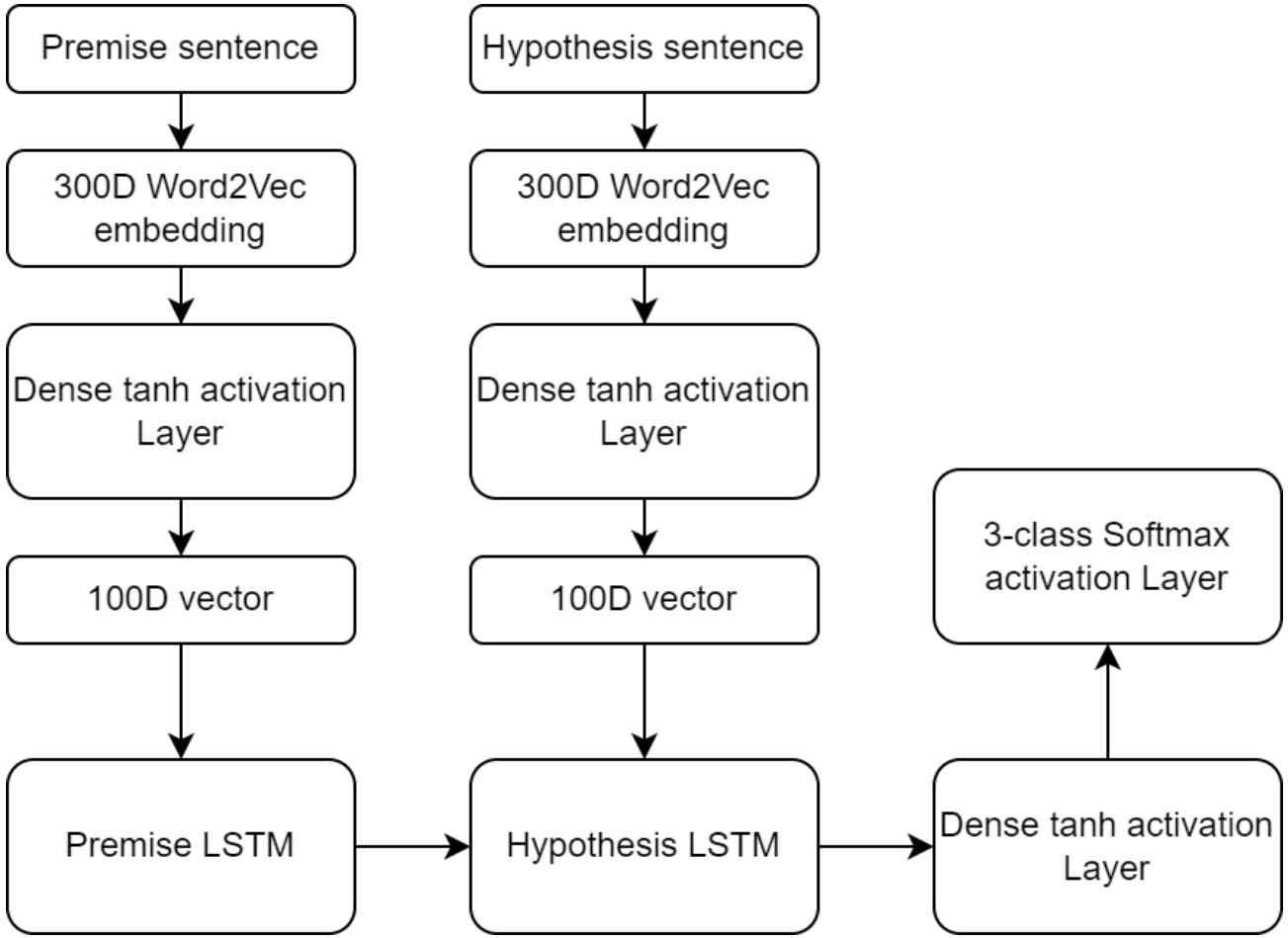


Results

	Training	Validation	Test
SNLI	75.8%	78.21%	77.3%
MultiNLI	67.41%	65.98%	65.76%
SciTail	83.77%	72.83%	74.58%
SICK	95.66%	73.14%	72.65%

Implementation 3: Two LSTMs Model

- Word2Vec used instead of GloVe.
- 300D embeddings sent to fully connected layer by tanh activation, output is 100D vectors.
- First LSTM inputs premise, returns final internal states.
- These states used to initialize second LSTM, which inputs hypothesis.
- 100D output of hypothesis LSTM fed into fully connected layer with tanh activation, followed by another fully connected layer with softmax over 3 labels.



Results

	Training	Validation	Test
SNLI	81%	80%	80.13%
MultiNLI	70.29%	68.65%	67.47%
SciTail	97.97%	70.53%	74.58%
SICK	98.60%	58.79%	60.57%

Test accuracies across models

	Summation	LSTM	2LSTM
SNLI	76%	77.3%	80.13%
MultiNLI	65.74%	65.76%	67.47%
SciTail	71.62%	74.58%	74.58%
SICK	70.49%	72.65%	60.57%

Thank You

Any Questions?