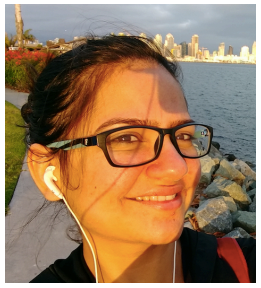
Relating Simple Sentence Representations in Deep Neural Networks and the Brain

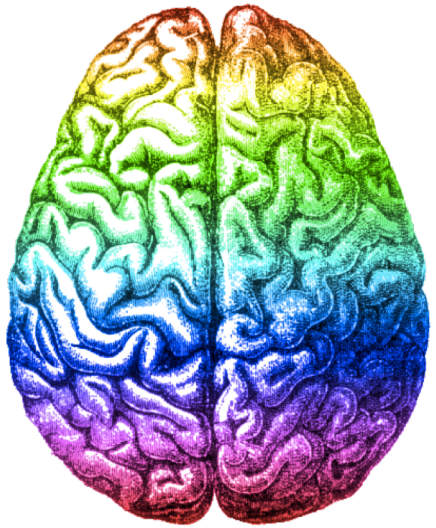


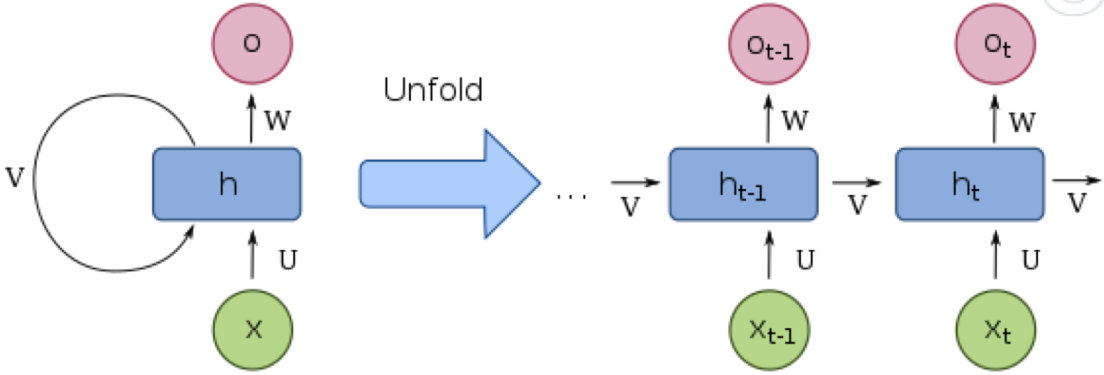

Sharmistha Jat[1] Hao Tang[2] Partha Talukdar[1] Tom Mitchell[2]

1- Indian Institute of Science, Bangalore
2- Carnegie Mellon University, Pittsburgh

Language representation in DNN and Brain

the dog ate the bone





Best language processing machine

Deep Neural Networks (DNN)

Images courtesy Wikipedia

Related Work

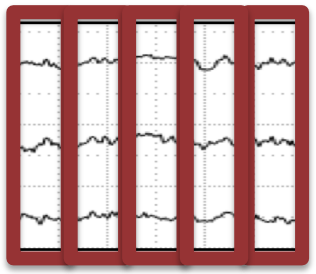
	DNN	**Main Findings**
Leila et al. EMNLP 2014	RNNLM (Mikolov 2012); NPLM (Vaswani 2013)	Parallelism between brain and DNN context representations
Our Paper	LSTM LM, ELMO, BERT, MULTITASK	Compare multiple DNN models vs the Brain; Detailed context evaluations; Brain data synthesis;

MEG Data

- Non-invasive method to record brain activity
- 306 sensors
- High temporal resolution: 1 milliseconds

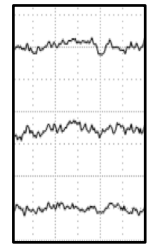
Word	*the*		*dog*		*ate*		*the*		*bone*	
Time(ms)	200	300	200	300	200	300	200	300	200	300

306 sensors



500 ms

Average 100ms data (non-overlapping)



306 sensors

5 points

Images courtesy Elekta.com

MEG Data Collection

- Record MEG data when human subjects read simple sentences
  - Active: *'the dog ate the bone'*
  - Passive: *'the bone was eaten by the dog'*

Dataset	# Sentences	Voice	Repetition
PassAct1	32	Passive+Active	10
PassAct2	32	Passive+Active	10
Act3	120	Active	10

Data and DNN Models

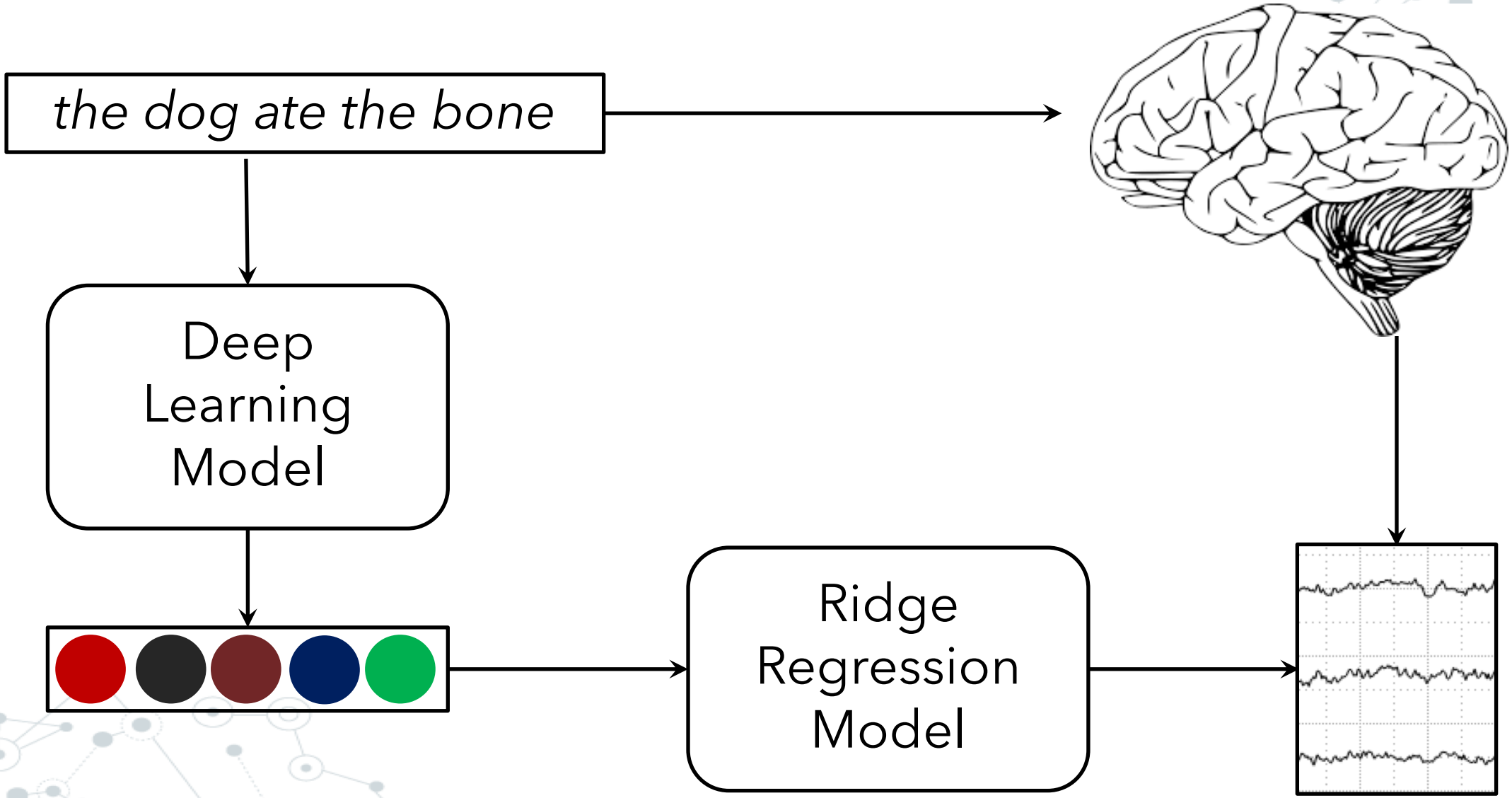
Simple Sentence Corpus (SSC):

Dataset	# Sentences	Voice
Wikipedia	125,900	Passive+Active
NELL Triples	130,245	Passive+Active

Models:

Random Embedding	Multi-task (LM & POS)
GloVe Additive	ELMO
Bi-LSTM LM	BERT

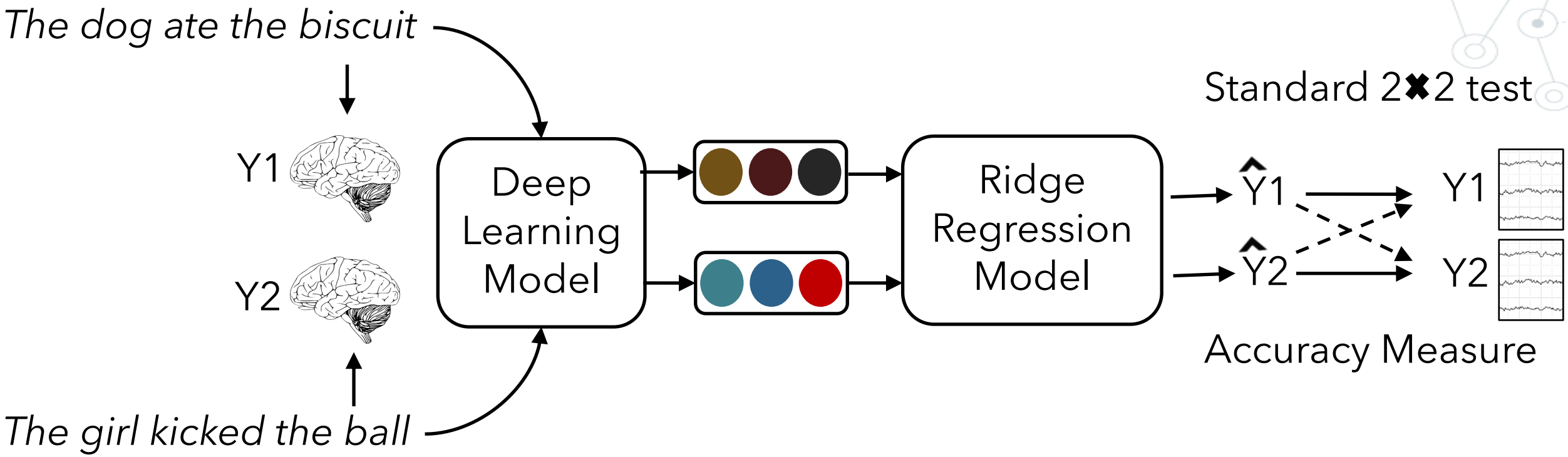
Experiment Setup



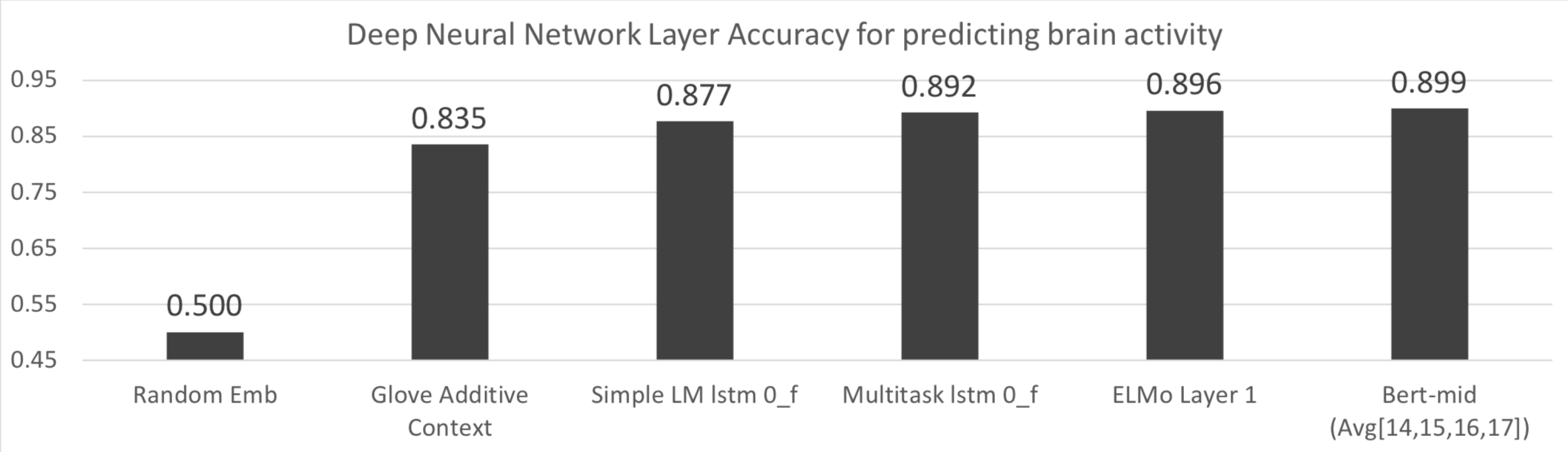
Experiments

- Relating the Brain and DNN representation:
  - Experiment 1 – Evaluating model's ability to predict brain activity accurately
  - Experiment 2 – Evaluating model's retention of past context using brain as reference
- Experiment 3 – Brain data augmentation using DNN

Experiment 1: Brain activity prediction

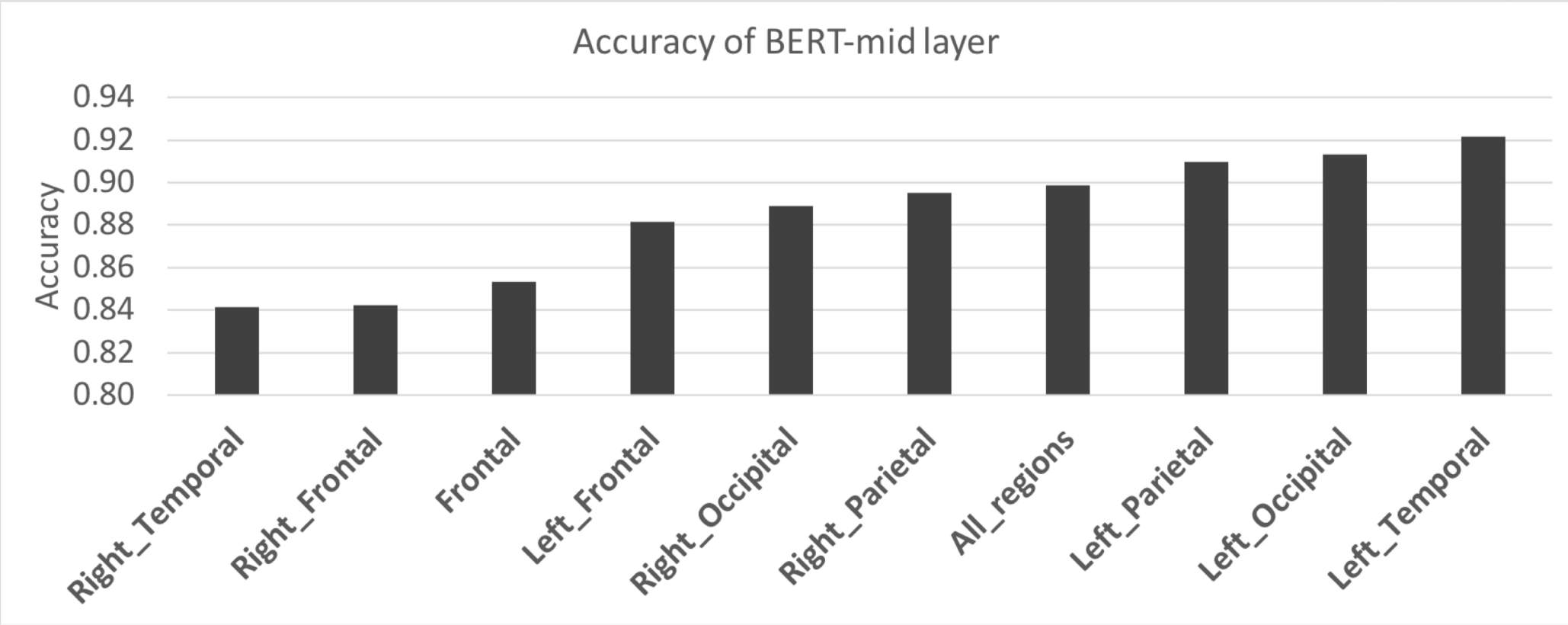


Experiment 1: Brain activity prediction



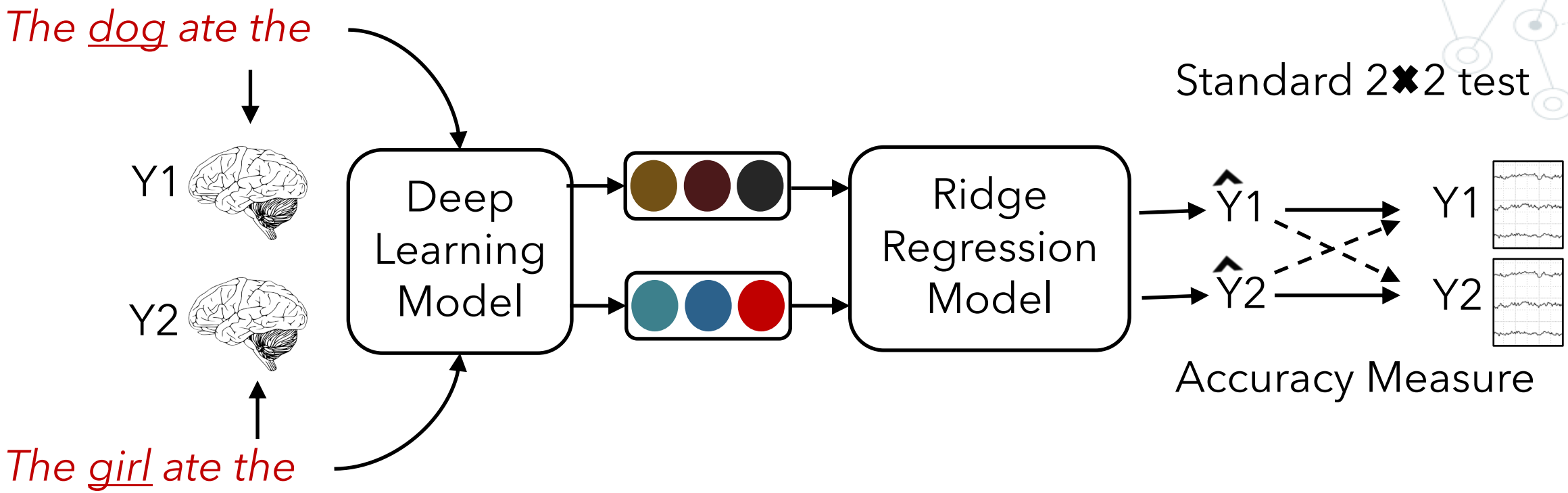
BERT-mid is most effective at predicting Brain activations

Experiment 1: Brain activity prediction



Left temporal brain region is predicted with highest accuracy

Experiment 2:



Experiment 2

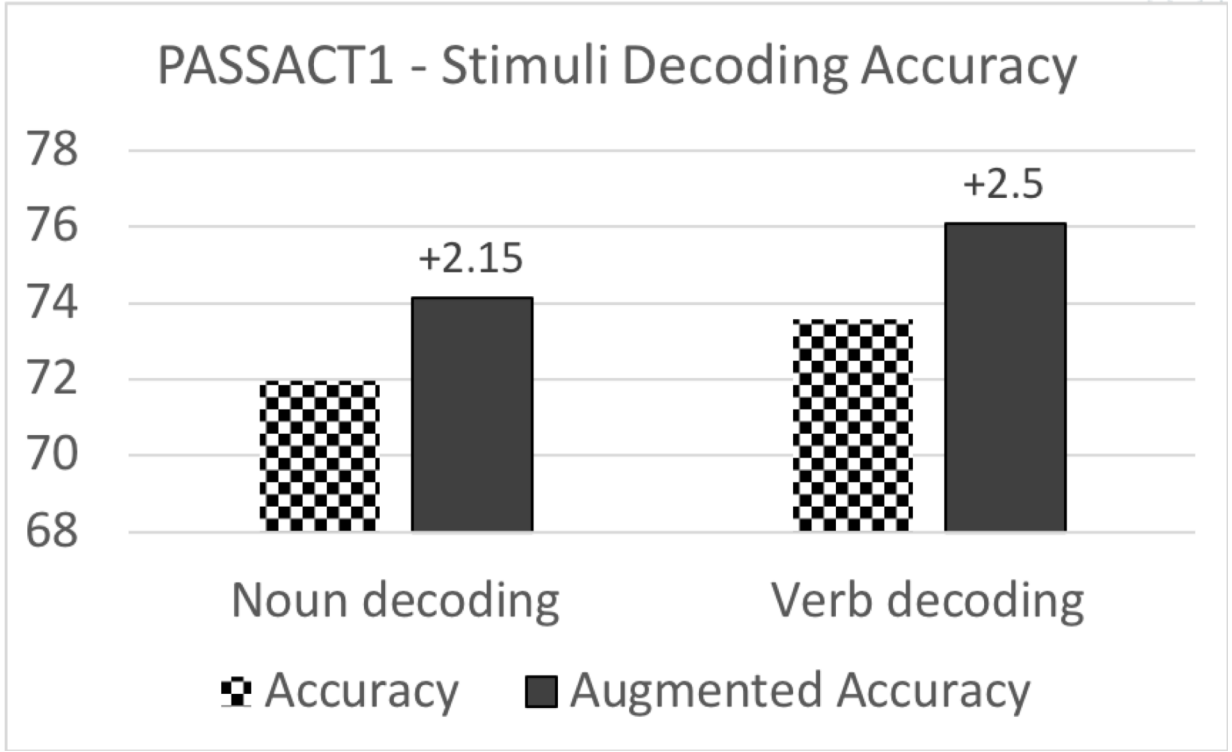
NOUN	"the dog ate the" vs "the girl ate the"	Most DNN layers retain Noun info	$ELMO_{mid}(0.92)$
VERB	"the dog ate the" vs "the dog saw the"	Most DNN layers retain Verb info	$ELMO_{mid}(0.92)$

Experiment 2

ADJECTIVE	"the happy child" vs "the child"	Middle DNN layers retain Adj info	Multitask LSTM layer1 (0.89)
FIRST DETERMINER	"the dog" vs "a dog"	Shallow DNN layers retain info better	BERT layer 3 (0.82) BERT layer 18 (0.78)

Experiment 3: Brain Data Augmentation

- Challenges:
  - High cost of collecting MEG recordings (~$60/sentence)
  - Subjects get fatigued inside the scanner quickly
- Augment brain data from BERT-mid representations



Conclusion

- Studied effect of prior context on DNN sequence representation and it's relationship with the Brain.
- BERT representations are the most predictive of Brain activity.
- BERT representations synthesize effective brain data for downstream data augmentation.

- Code link : bit.ly/2Ynx7Ek
- Paper link: arxiv.org/abs/1906.11861
- Contact : sharmisthaj@iisc.ac.in

Supported by:

PRATIKSHA TRUST

MHRD
GOVERNMENT OF INDIA

BrainHub
Carnegie Mellon University