UNIVERSITY AT ALBANY
STATE UNIVERSITY OF NEW YORK

IECE – 566 Deep Learning Presentation

CSI-Based Sign Language Recognition Using a CNN-GRU Architecture Enhanced with Attention

Joy Saha
Department of Electrical and Computer Engineering

Date: 12/02/25

AGENDA

Motivation and Problem Statement

Objective

CSI Overview

Propose Architecture

Experiments and Results

Conclusion

MOTIVATION AND PROBLEM STATEMENT

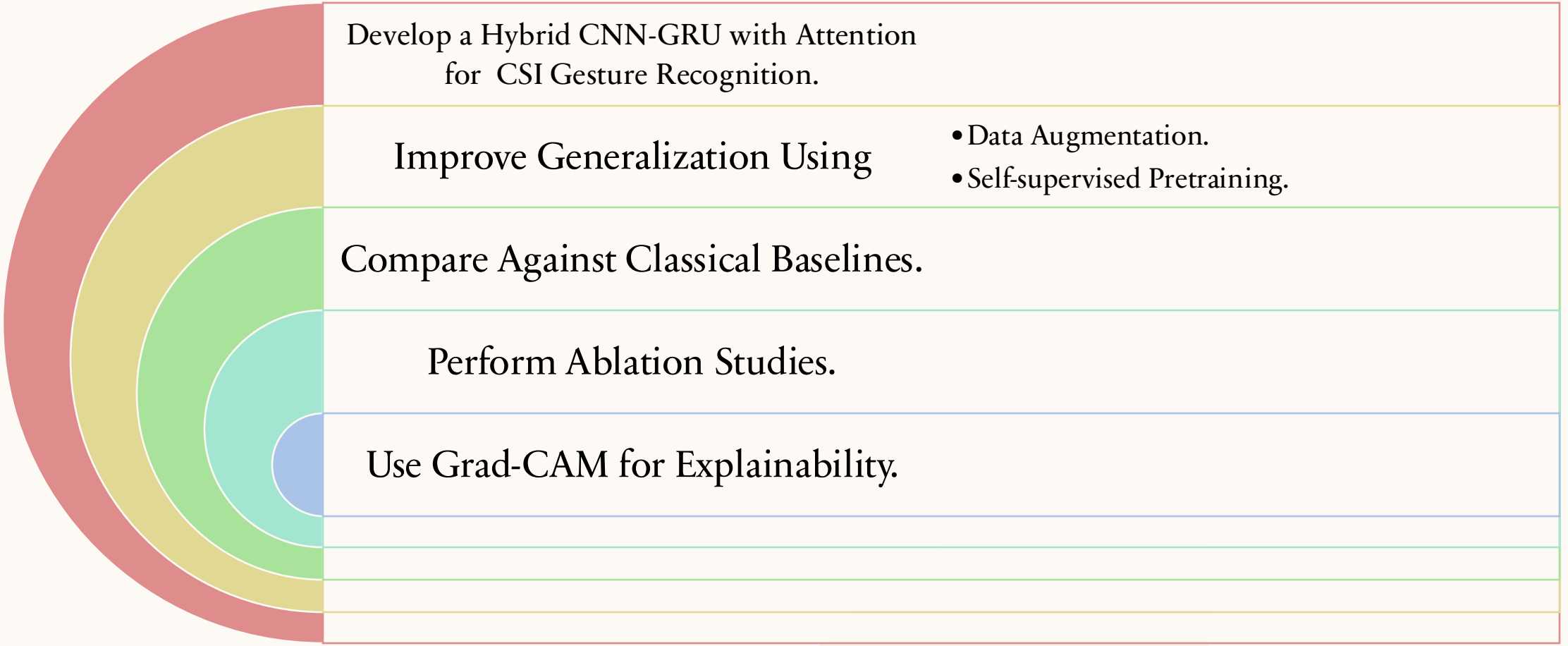
Sign Language Recognition Improves Communication Accessibility.

Challenges:

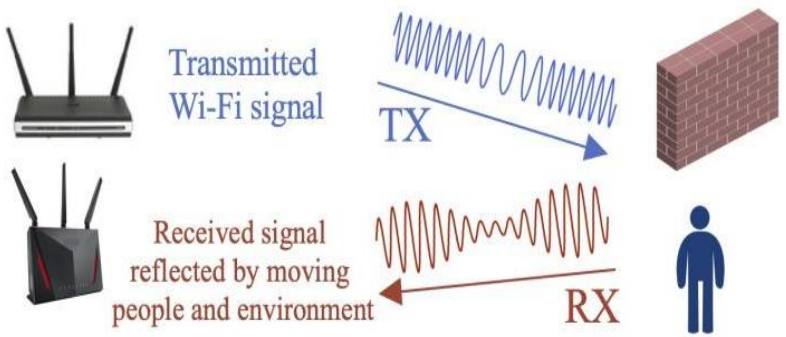
- High-dimensional CSI signals (amplitude & phase)
- Complex Spatial-temporal Patterns.

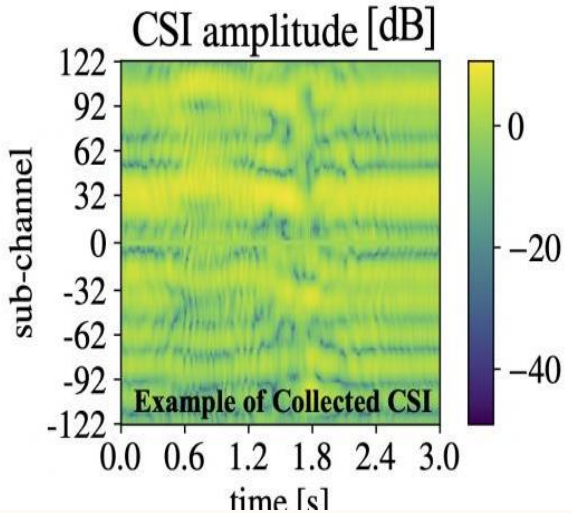
Wi-fi CSI: Privacy-preserving, Contactless Sensing.

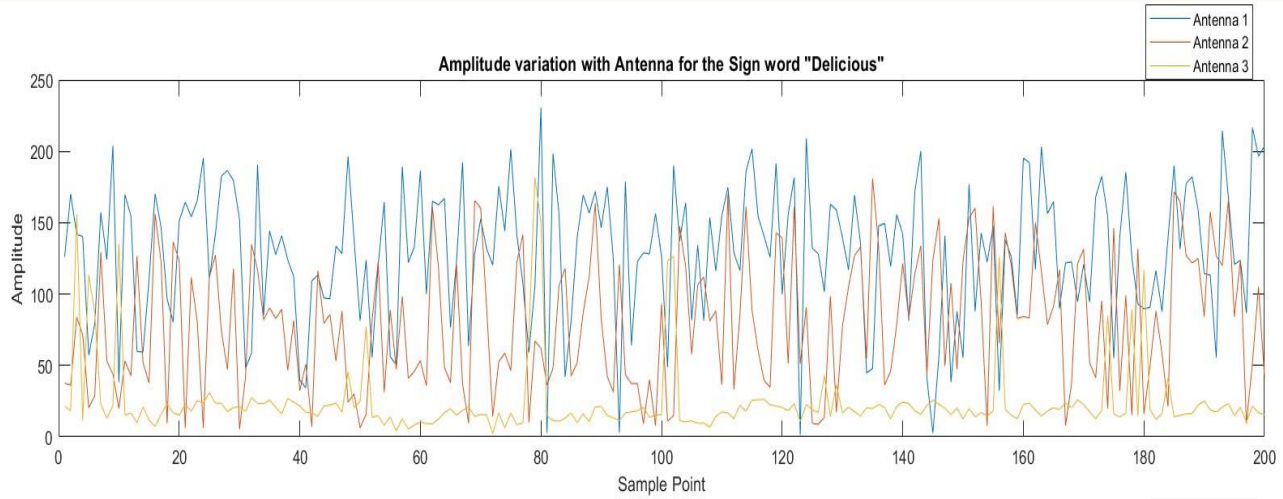
OBJECTIVE

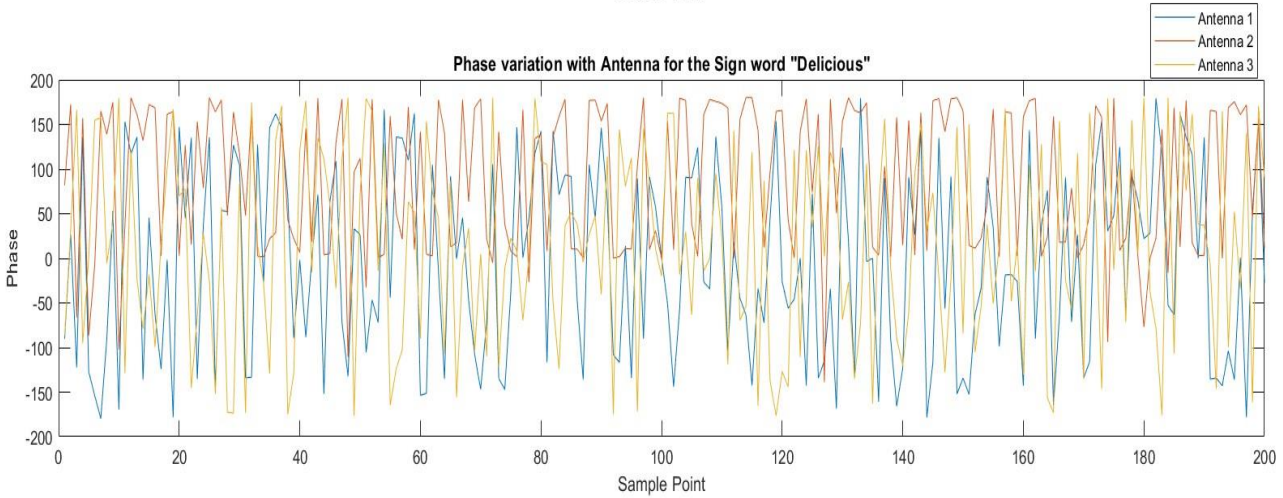

Develop a Hybrid CNN-GRU with Attention for CSI Gesture Recognition.
Improve Generalization Using
• Data Augmentation.
• Self-supervised Pretraining.
Compare Against Classical Baselines.
Perform Ablation Studies.
Use Grad-CAM for Explainability.

CSI OVERVIEW

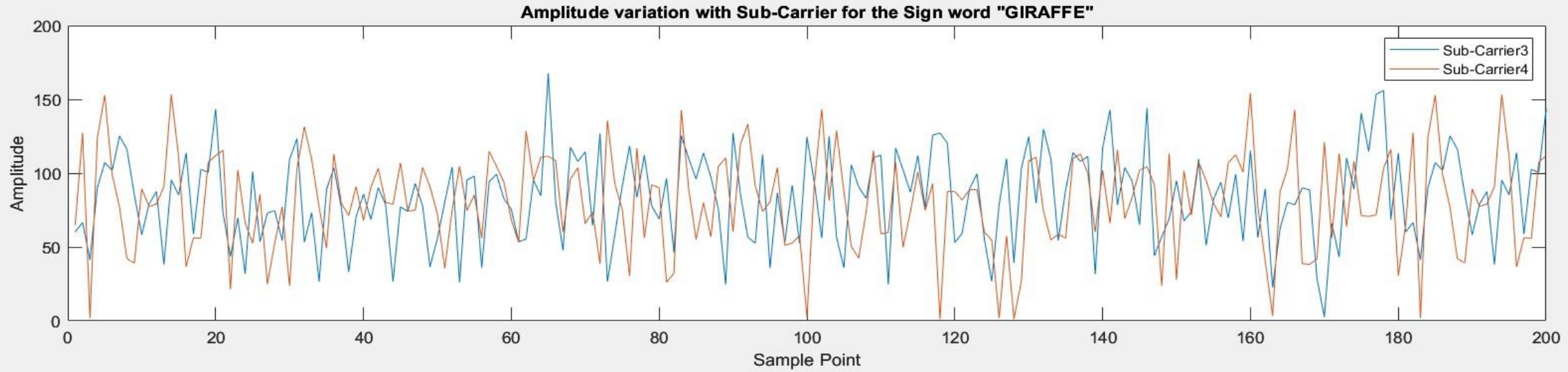


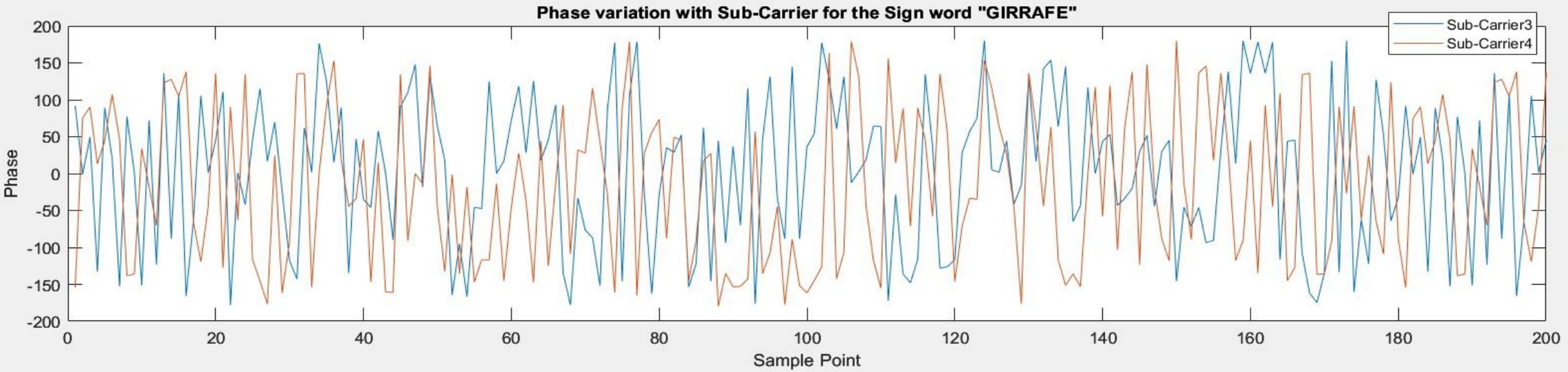




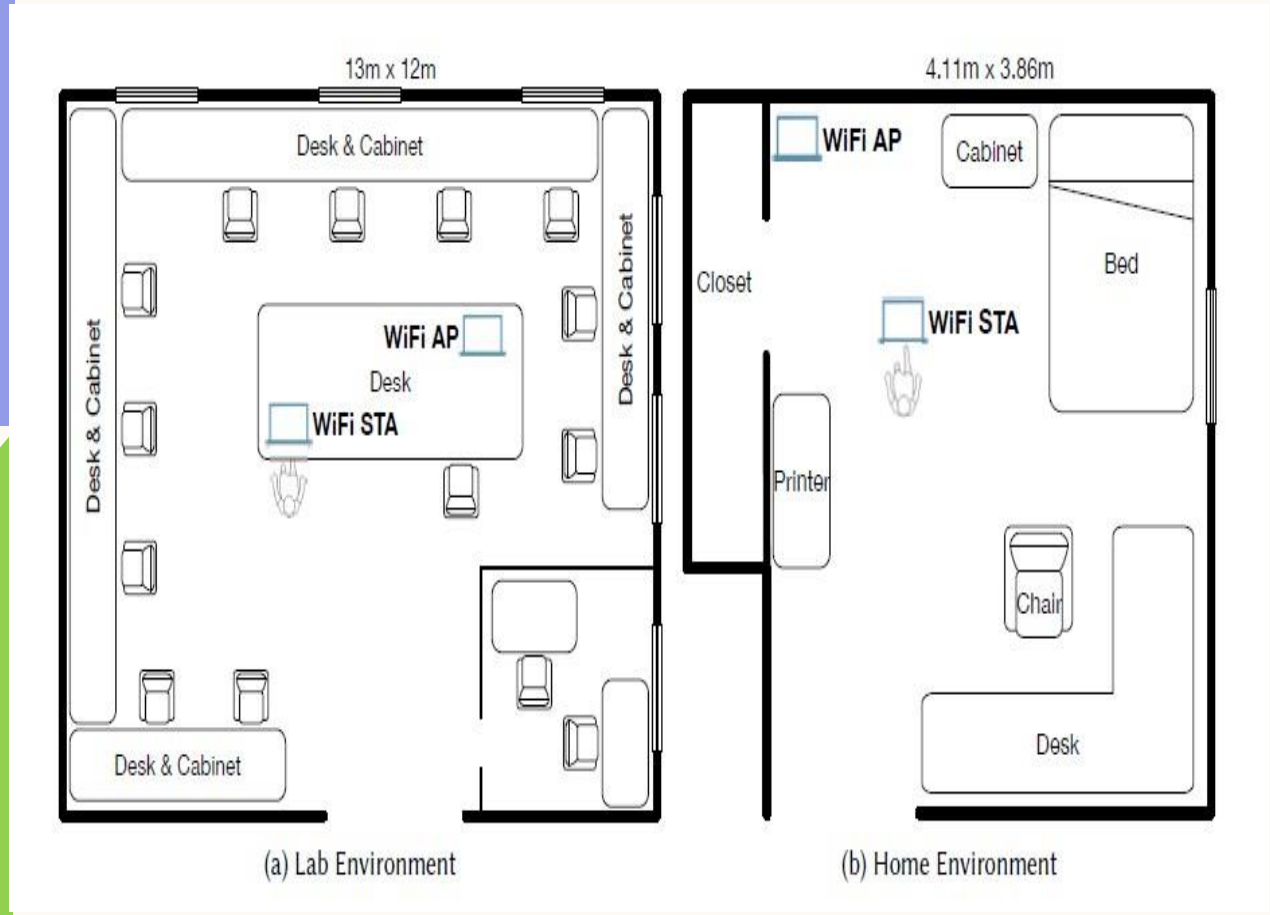


CSI OVERVIEW (CONT.)



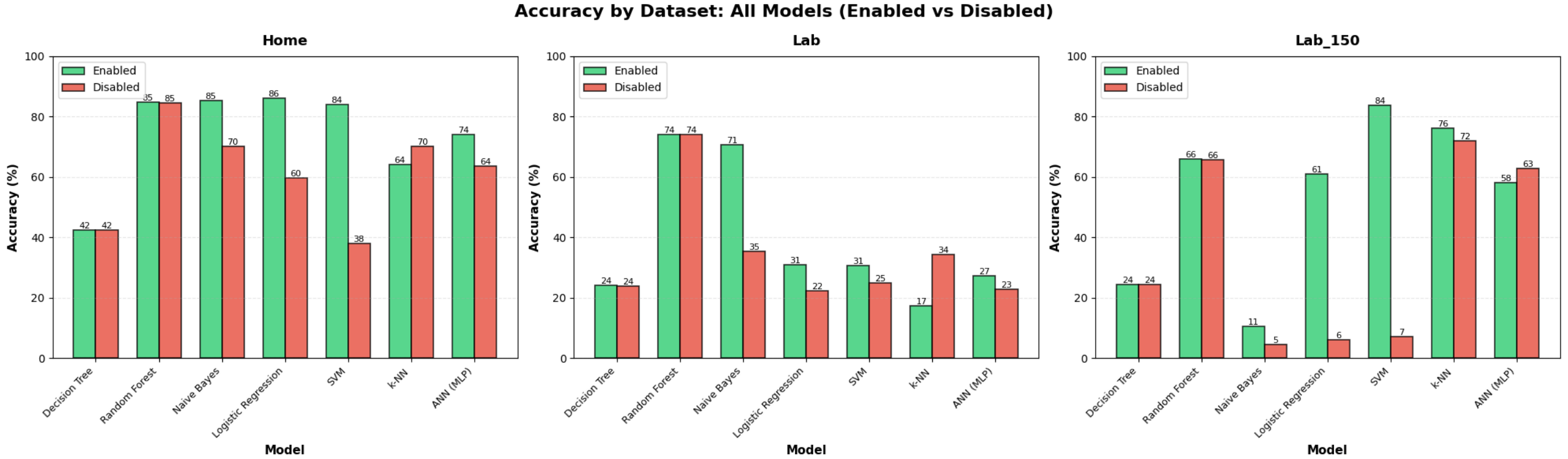


DATASET DETAILS



Dataset	# Labels	# Repetitions	# Instances
Home	276	10	2760
Lab	276	20	5520
Lab150	150	10	7500

BASELINE RESULTS

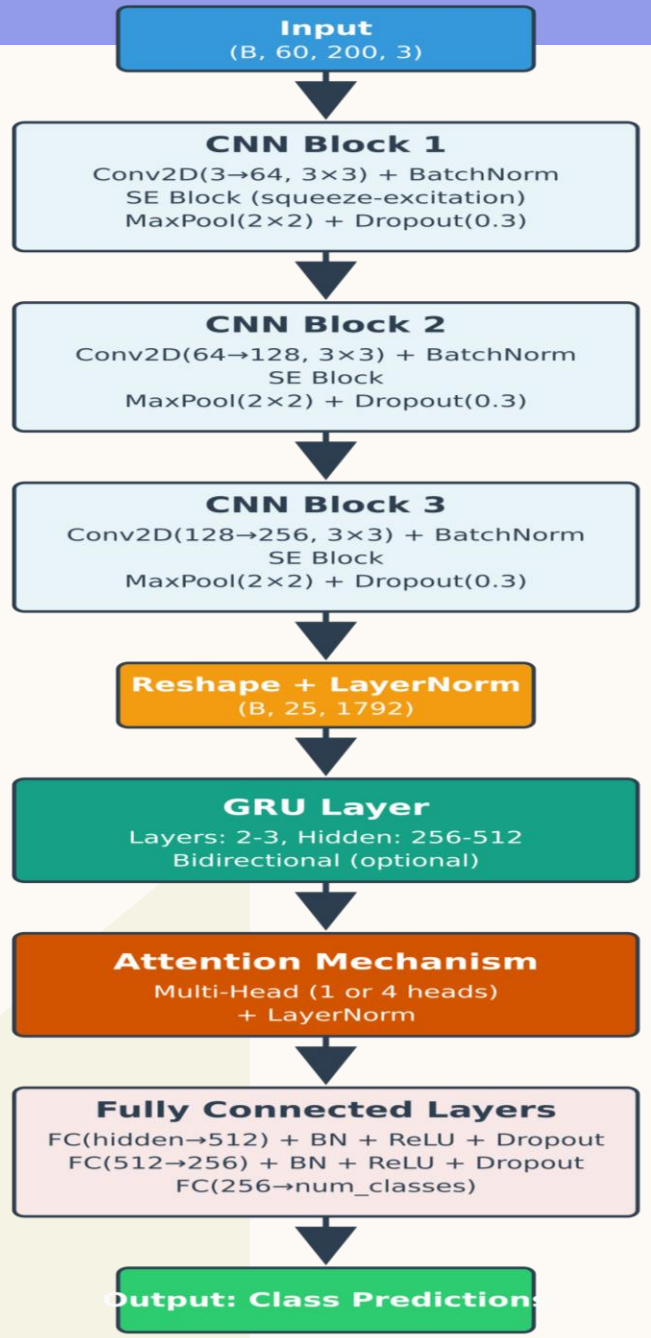


PROPOSED ARCHITECTURE

CNN: Extract Spatial Features from CSI.

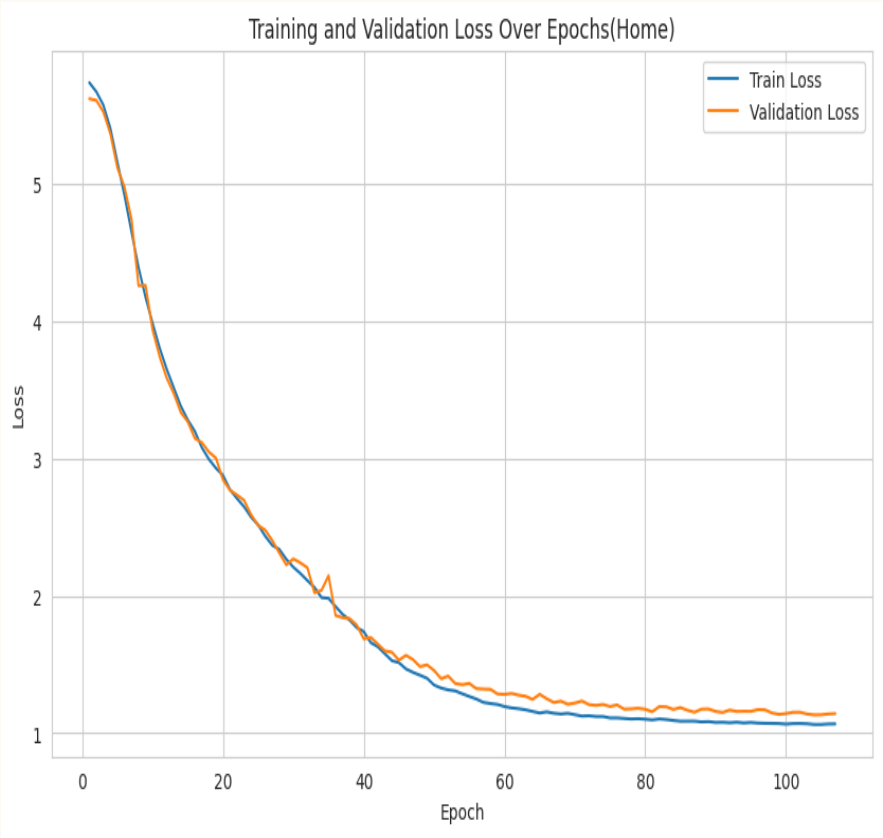
GRU: Capture Temporal dynamics.

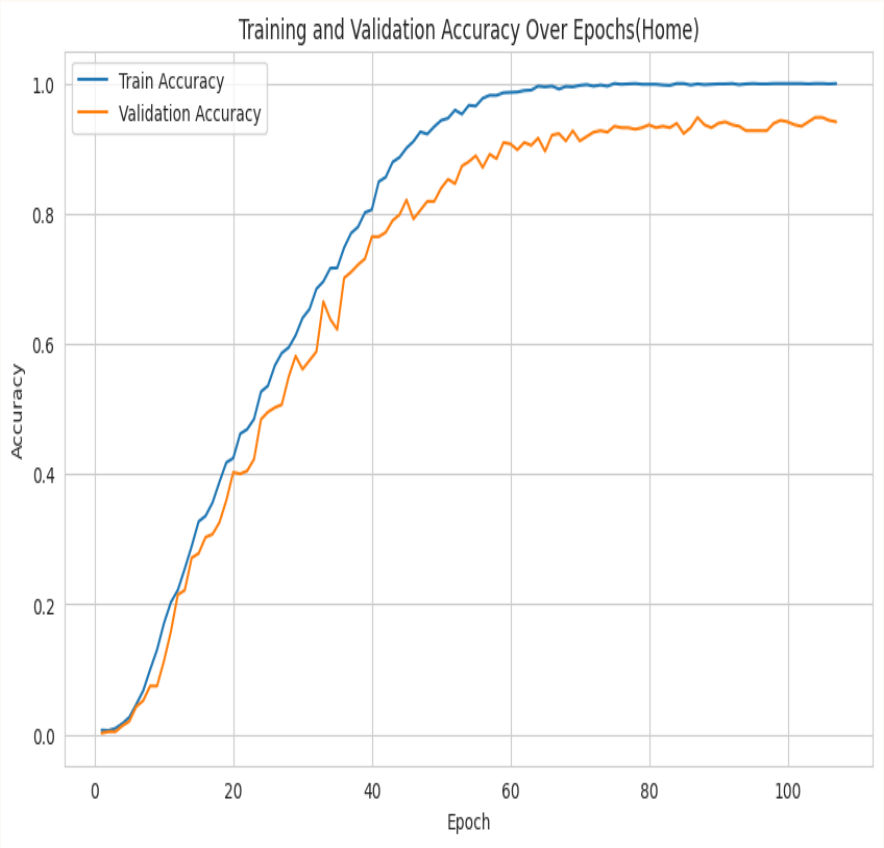
Attention: Focus on Salient Sequences.



EXPERIMENTS AND RESULTS

CNN GRU Attention





Dataset	Test Acc
Home	94.75%
Lab	93.93%
Lab 150	86.33%

EXPERIMENTS AND RESULTS (CONT.)

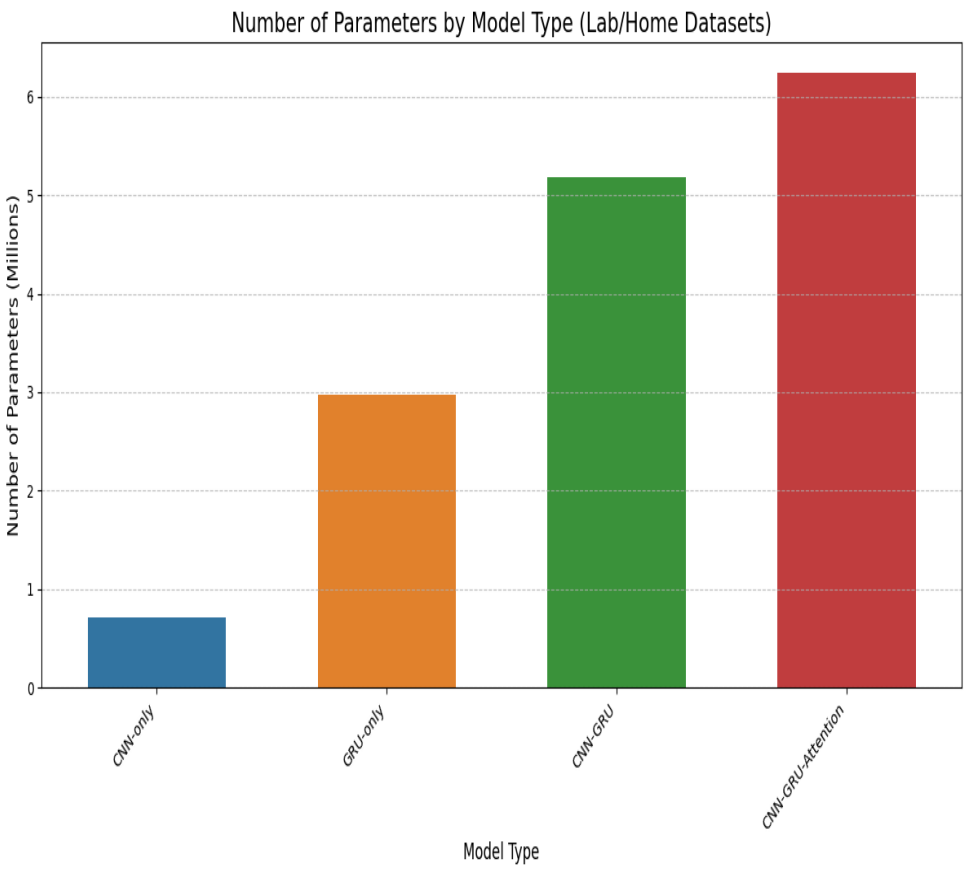
Effects of Multi-head Attention

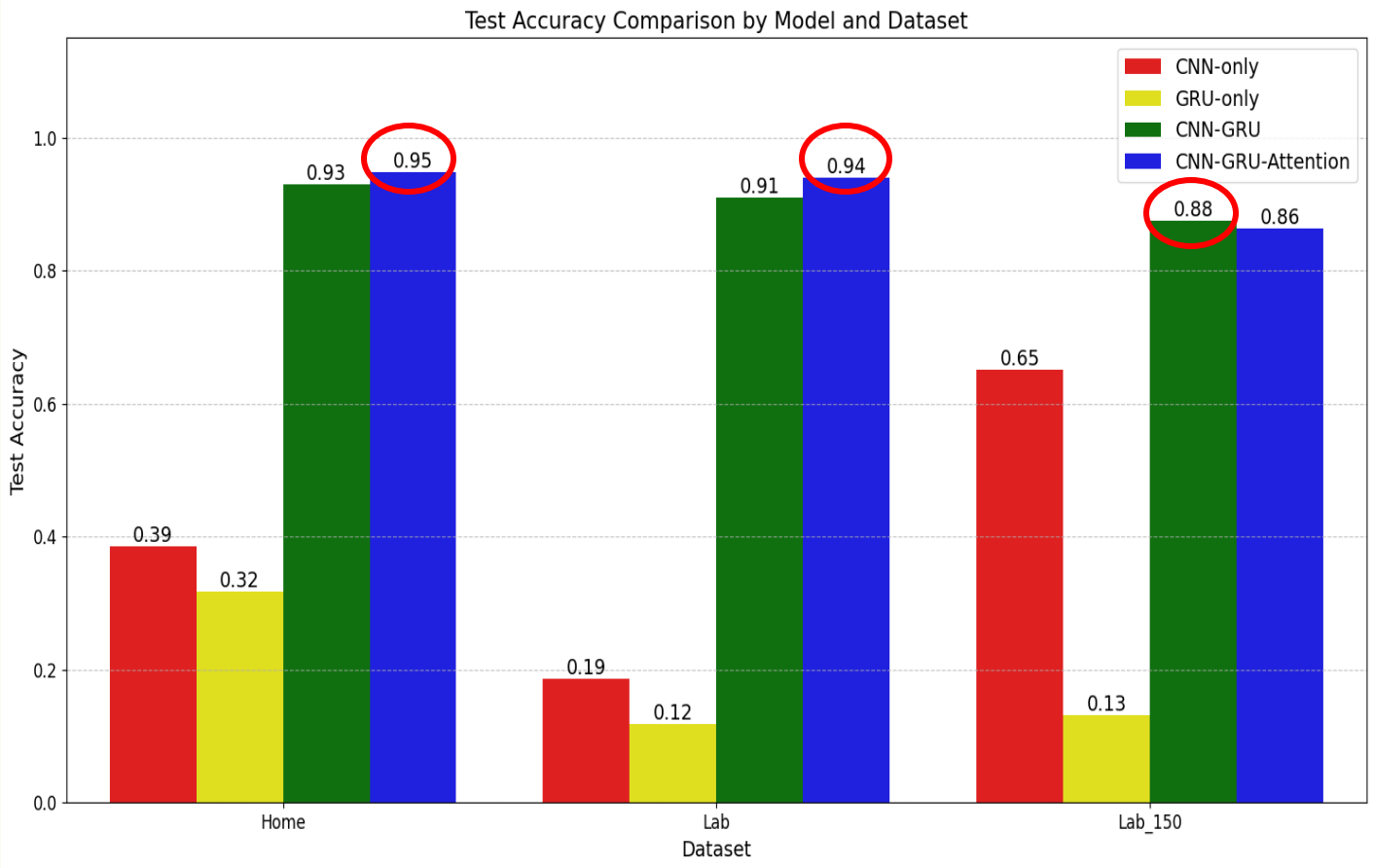
Dataset	1	2	4	8
Home	76.27%	90.76%	**94.75%**	86.78%
Lab	80.80%	**94.84%**	93.93%	95.11%
Lab_150	85.35%	84.93%	**86.33%**	84.27%

Effect of Bidirectionality

Dataset	True	False
Home	**94.75%**	89.31%
Lab	**93.93%**	91.76%
Lab_150	86.33%	**86.40%**

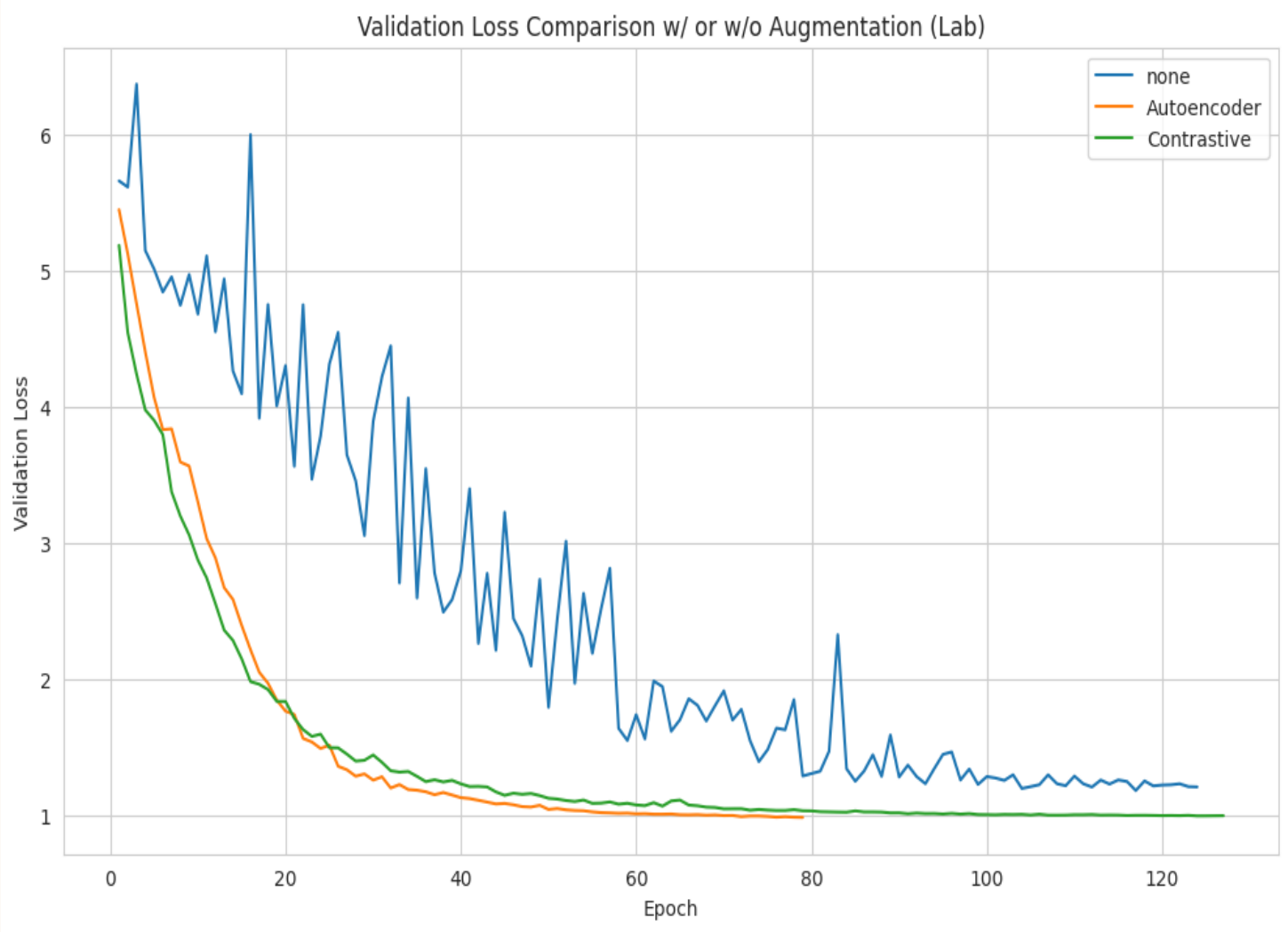
ABLATION STUDY



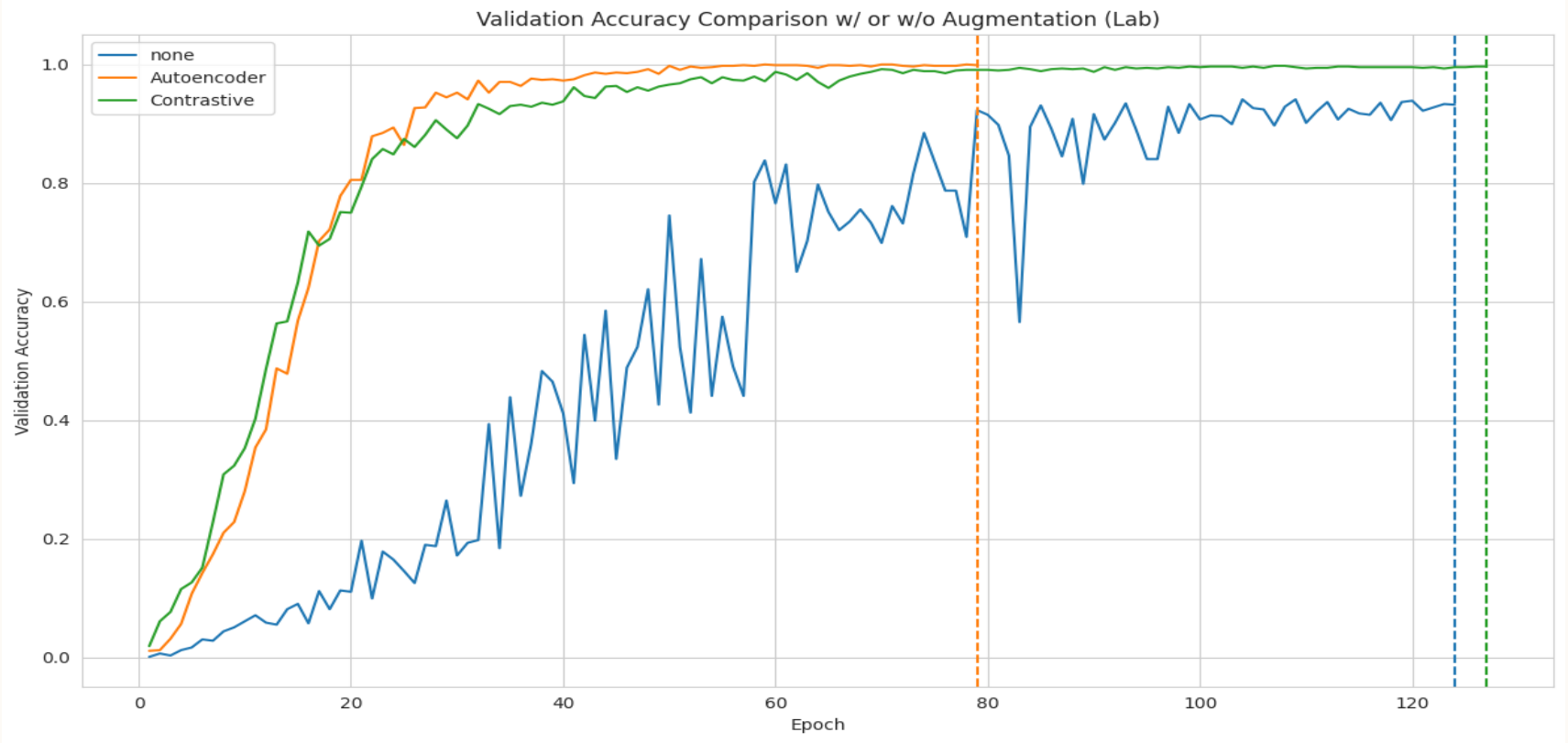


AUGMENTATION AND PRETRAINING

- Self-Supervised Pretraining
  - Auto-encoder.
  - Contrastive.
- Augmentation
  - Temporal Warping.
  - Gaussian Jitter.
  - Random Mask.
  - Wrap + Jitter.

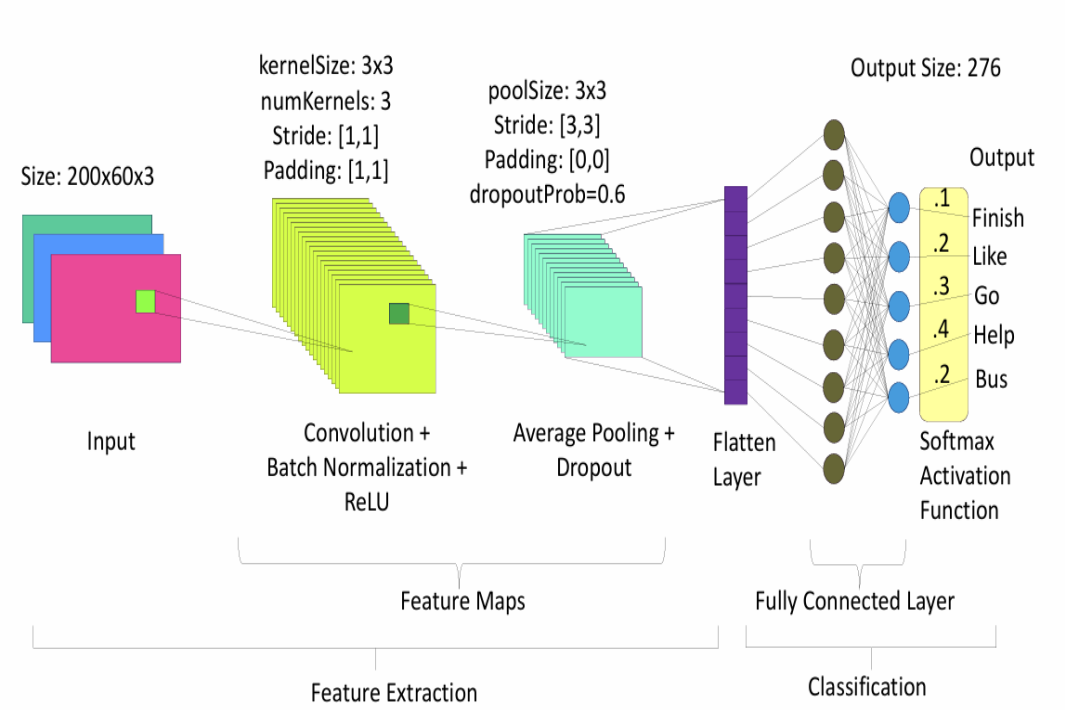


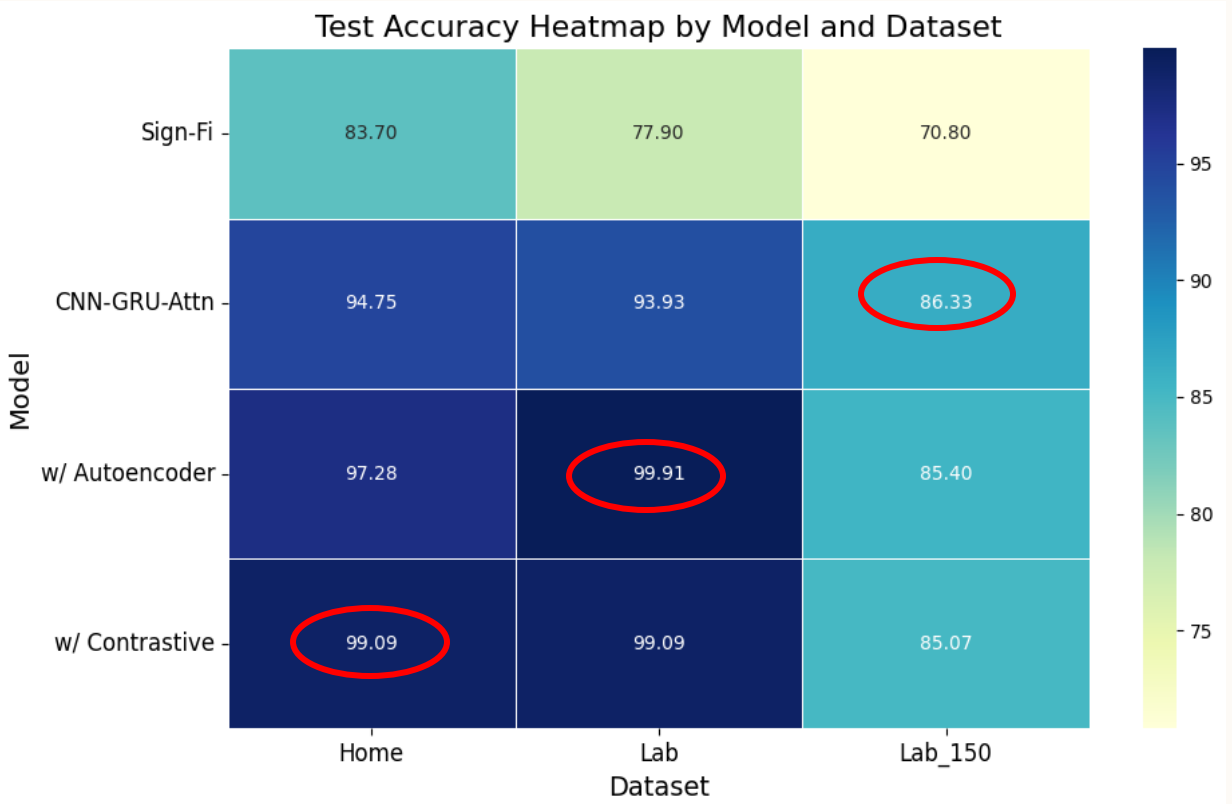
AUGMENTATION AND PRETRAINING (CONT.)



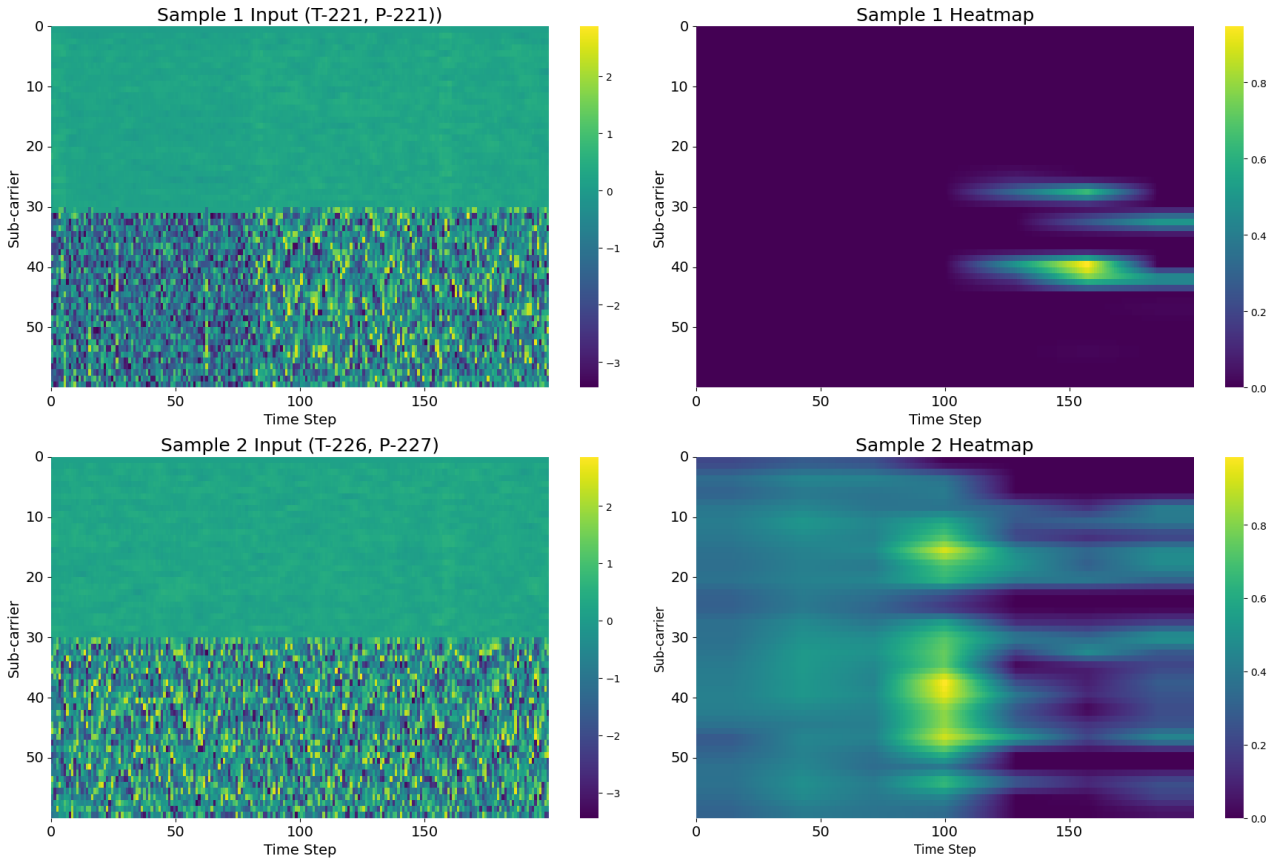
Dataset	None	Auto-encoder	Contrastive
Home	94.75%	97.28%	**99.09%**
Lab	93.93%	**99.91%**	99.09%
Lab_150	**86.33%**	85.40%	85.07%

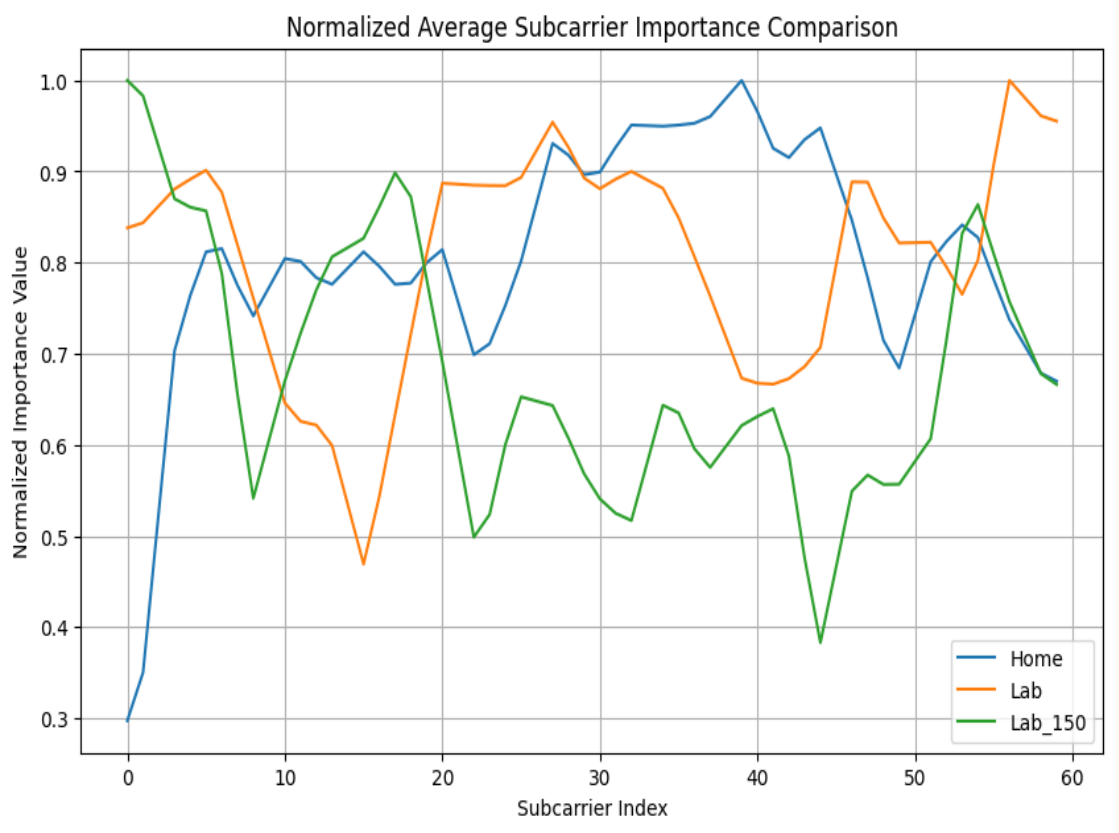
SIGN-FI COMPARISON





EXPLAINABILITY





LIMITATIONS

- High Computational Complexity.
- Fixed Input Shape.
- Limited Attention Flexibility.
- Dataset Dependence.

Held Out Test

Dataset	Train Acc	Test Acc
Home	99.16%	85.56%
Lab	99.54%	88.03%
Lab-150	99.00%	79.72%

CONCLUSION

CNN-GRU with Attention Effectively Classifies CSI-based Gestures.

Potential for Broader Human Activity Sensing.

Future Work: Larger Datasets, Real-time deployment.

More Robust Model to Include Limited / Noisy CSI data.

REFERENCES

- Y. Ma, G. Zhou, S. Wang, H. Zhao, W. Jung, “Sign-Fi: Sign Language Recognition Using Wi-Fi,” Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 2,no. 1, pp. 1–21, Mar. 2018.
- Francesca Meneghello, Nicolò Dal Fabbro, Domenico Garlisi, Ilenia Tinnirello, Michele Rossi, "CSI Dataset for Wireless Human Sensing on 80 MHz Wi-Fi Channels", IEEE Dataport, October 30, 2022, doi:10.21227/xbhv-f125.
- http://www.startasl.com/asl-dictionary/

Thank you !

Any Question ?