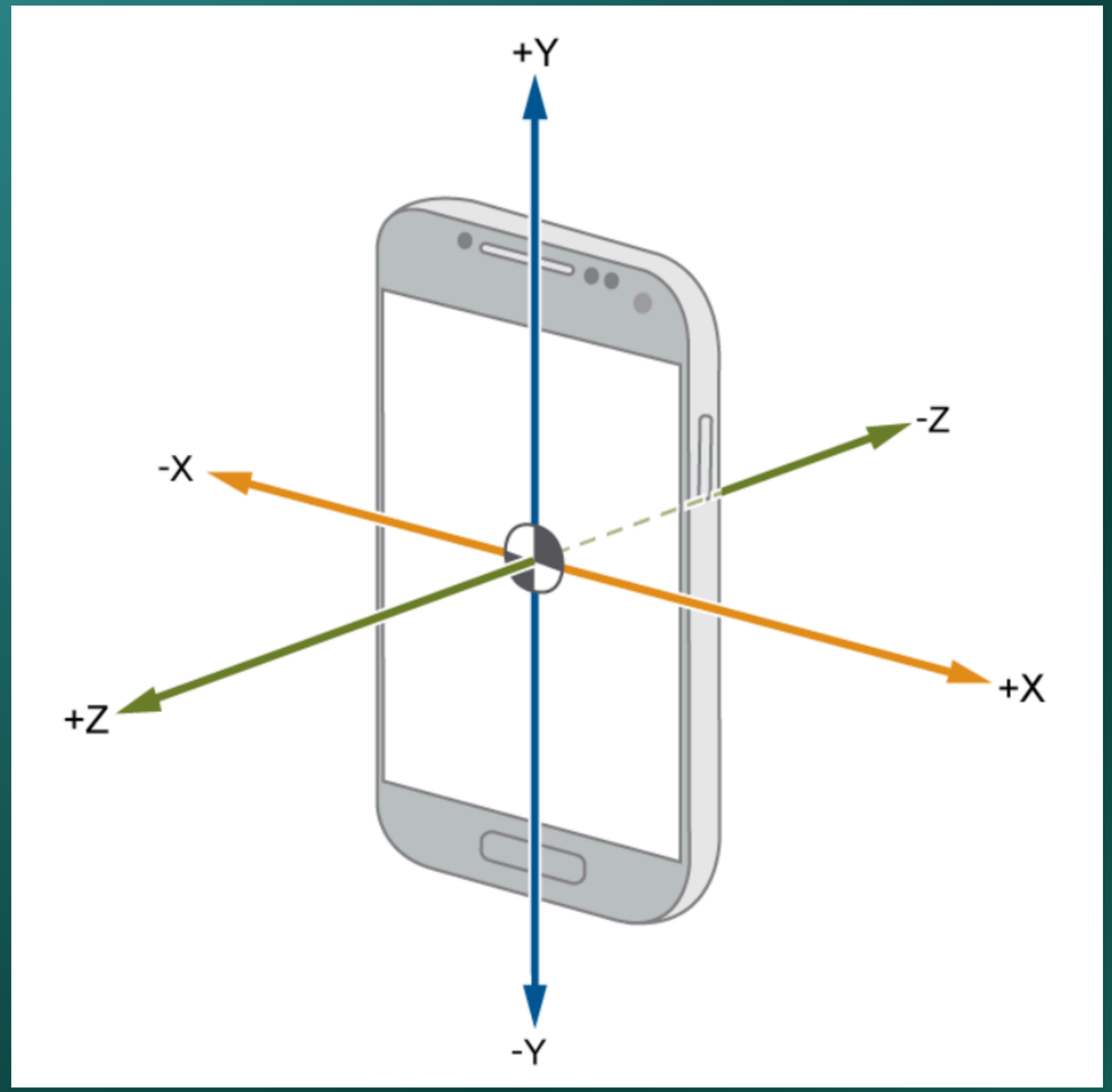
Human Activity Recognition (HAR)

Using smartphone sensor data to teach machines how we move

Neural Network SS 25

Mohan Adluru & Vanessa Geiss



- Accelerometers & gyroscopes track motion along X, Y, Z
- Raw sensor values reflect user activity
- Patterns in this data allow us to predict behaviors
- Basis for AI models in Human Activity Recognition

Human Activity Recognition with Smartphone sensory

UCI Human Activity Recognition (HAR) Dataset

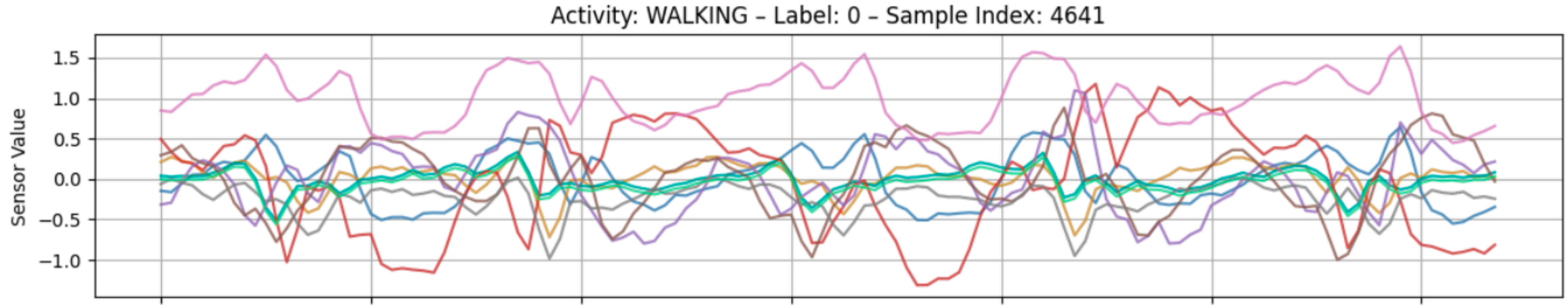
10,299 samples from 30 participants
Smartphone worn on waist
Sensors: Accelerometer + Gyroscope
Sampling rate: 50 Hz

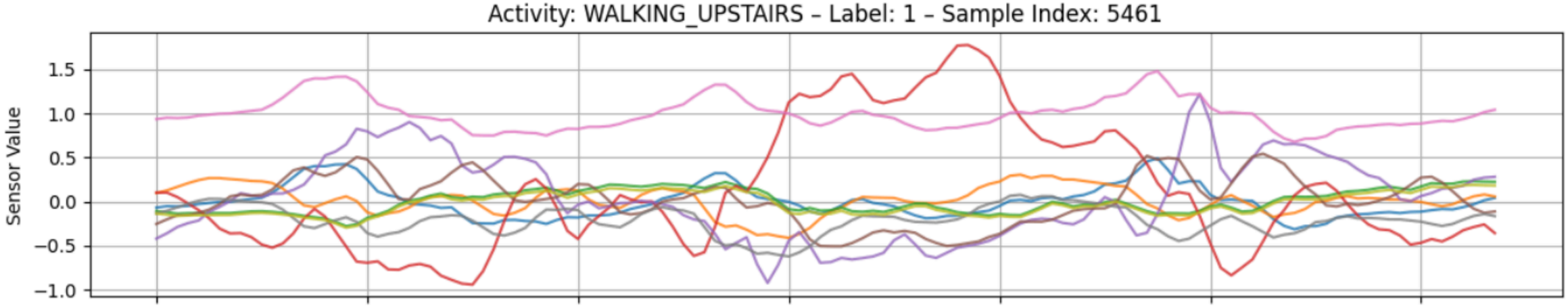
6 Activities:

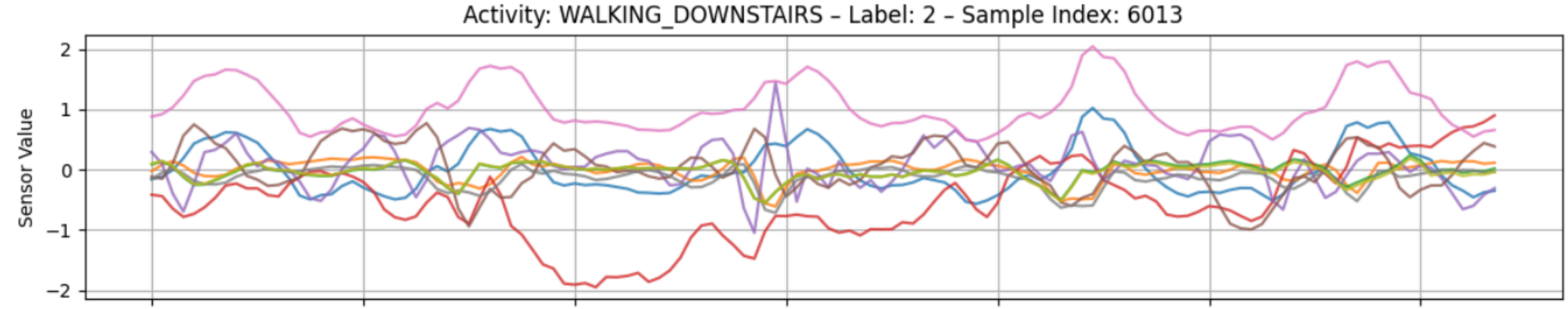
- Walking
- Walking Upstairs
- Walking Downstairs
- Sitting
- Standing
- Laying

Format: 128 time steps × 9 features per sample

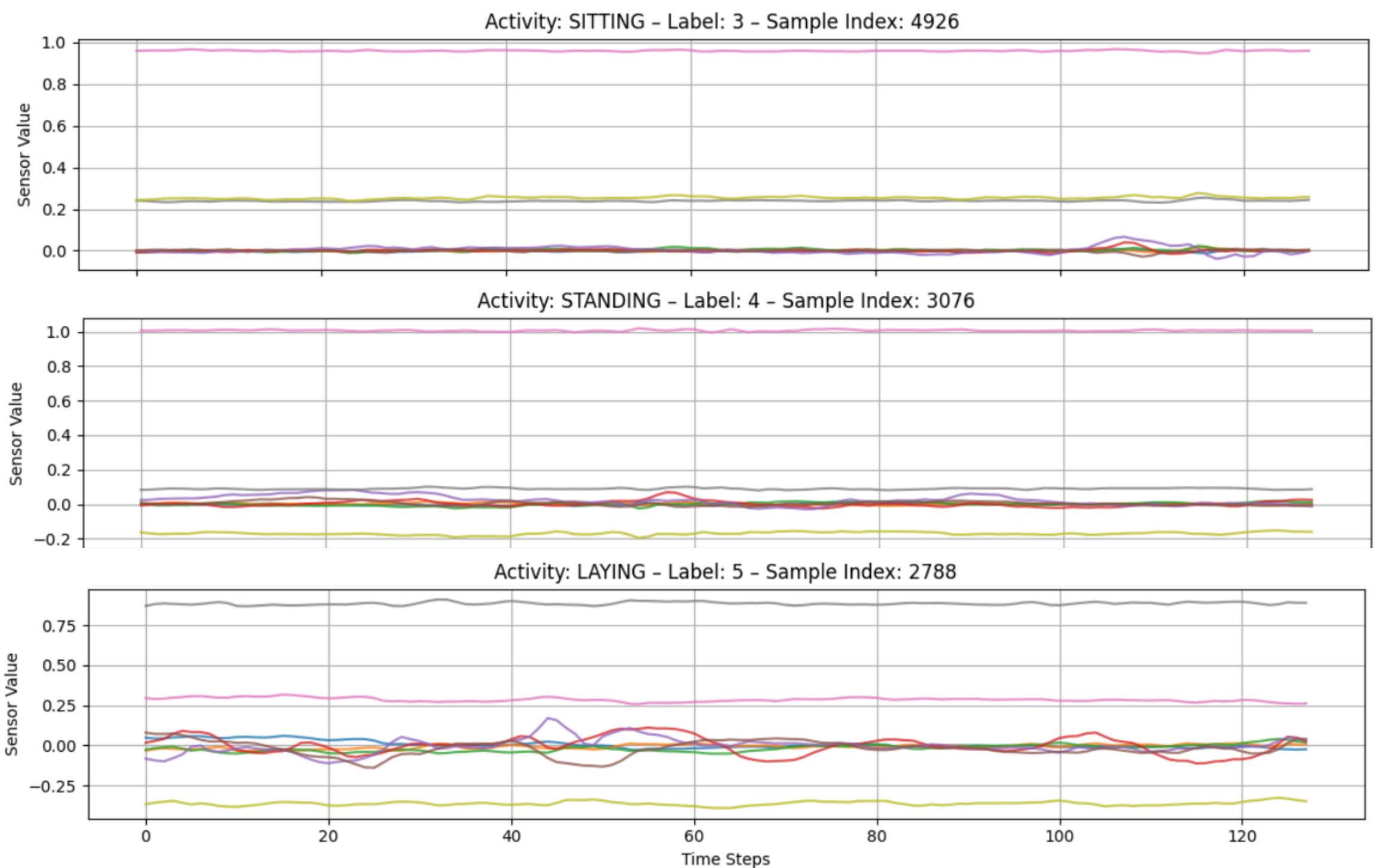
Activity


Activity: WALKING – Label: 0 – Sample Index: 4641
Sensor Value
1.5
1.0
0.5
0.0
−0.5
−1.0


Activity: WALKING_UPSTAIRS – Label: 1 – Sample Index: 5461
Sensor Value
1.5
1.0
0.5
0.0
−0.5
−1.0


Activity: WALKING_DOWNSTAIRS – Label: 2 – Sample Index: 6013
Sensor Value
2
1
0
−1
−2

Activity


Activity: SITTING – Label: 3 – Sample Index: 4926
Sensor Value
1.0
0.8
0.6
0.4
0.2
0.0
Activity: STANDING – Label: 4 – Sample Index: 3076
Sensor Value
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
Activity: LAYING – Label: 5 – Sample Index: 2788
Sensor Value
0.75
0.50
0.25
0.00
−0.25
0
20
40
60
80
100
120
Time Steps

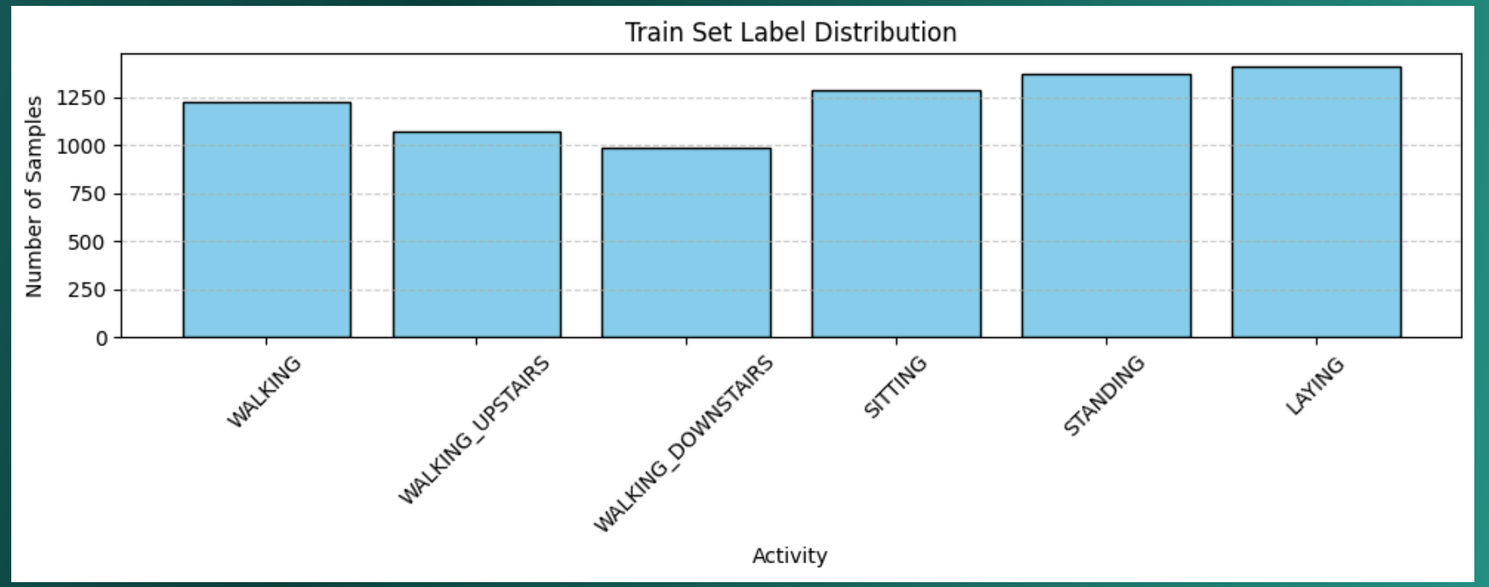
Labels

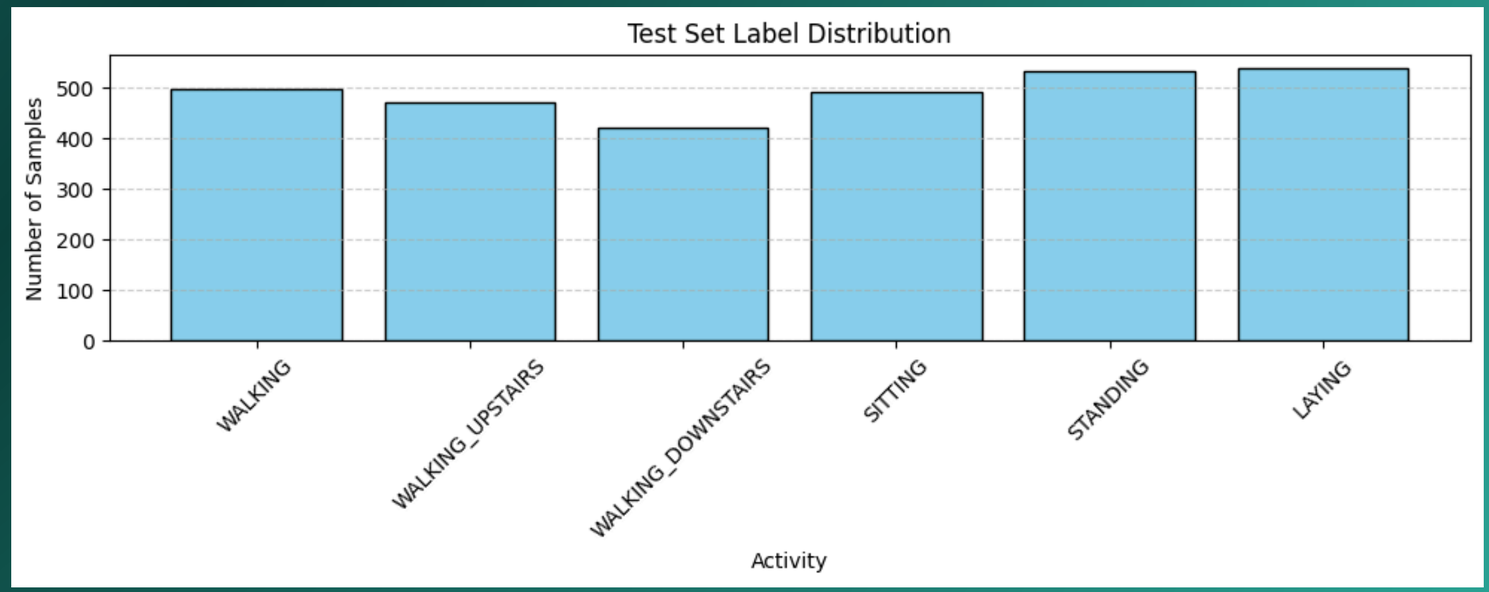
Label Distribution in **TRAIN** Set:

Label 0: 1226 samples (16.68%)
Label 1: 1073 samples (14.59%)
Label 2: 986 samples (13.41%)
Label 3: 1286 samples (17.49%)
Label 4: 1374 samples (18.69%)
Label 5: 1407 samples (19.14%)

Label Distribution in **TEST** Set:

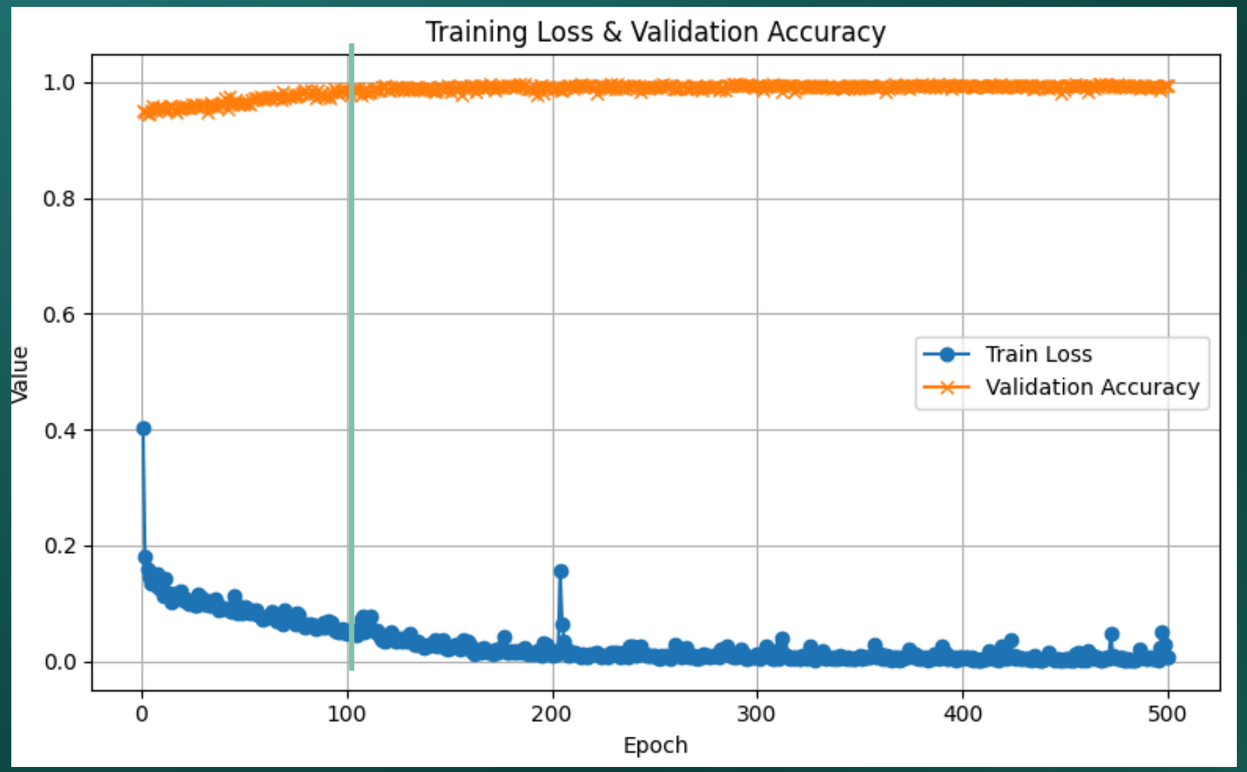
Label 0: 496 samples (16.83%)
Label 1: 471 samples (15.98%)
Label 2: 420 samples (14.25%)
Label 3: 491 samples (16.66%)
Label 4: 532 samples (18.05%)
Label 5: 537 samples (18.22%)





What it Learns

- Local temporal patterns in short windows
- Fast, efficient, great at extracting motion spikes
- Misses longer temporal relationships



1D CNN Architecture

Input:
9 × 128

↓

Conv1D (64 filters, k=3) + ReLU
+ MaxPool1D (k=2) Output: 64 × 64

↓

Conv1D (256 filters, k=3) + ReLU
+ MaxPool1D (k=2) Output: 128 × 32

↓

Conv1D (256 filters, k=3) + ReLU
+ MaxPool1D (k=2) Output: 256 × 16

↓

Flatten
D56 × I → 4086

+ ReLU
+ Dropout (0.5)

↓

Dense (6)
+ RELU

Logits
6 Activity Classe

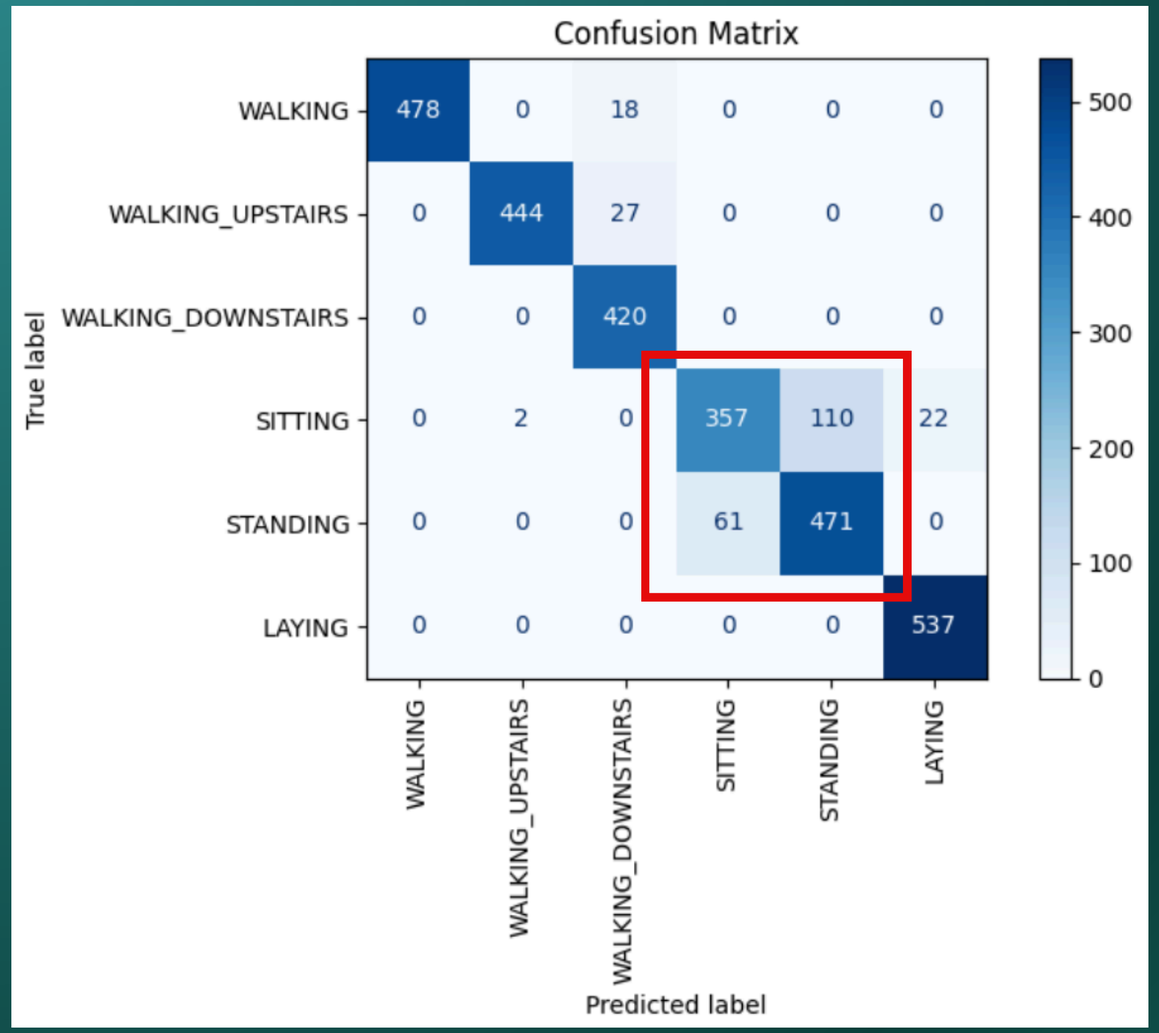
Optimizer:
Adam (lr=0.001)

Loss Function:
CrossEntropyLoss

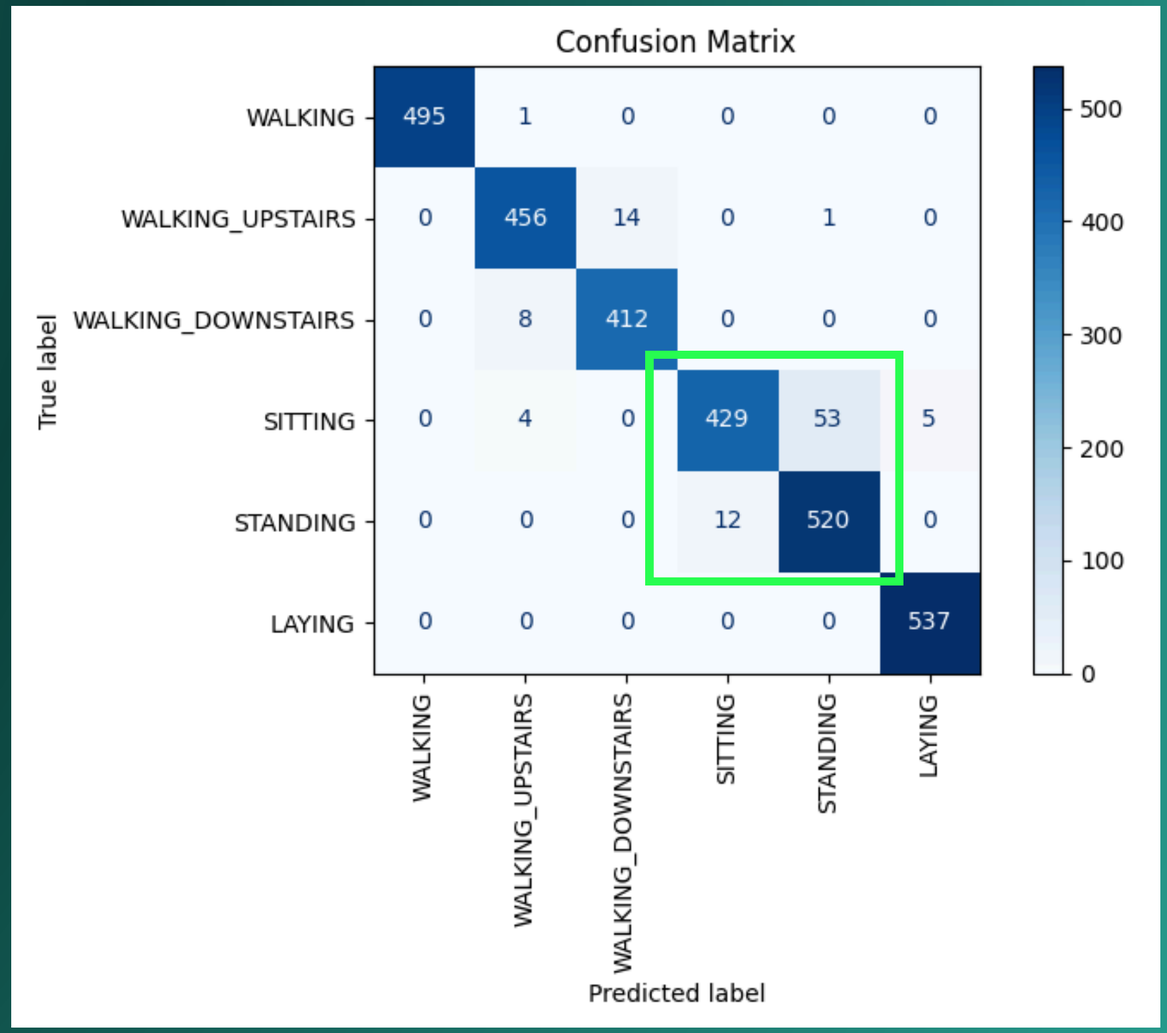
Dropout: 0.5

Input Shape:
(Batch, 9, 128)

Output: 6 classes

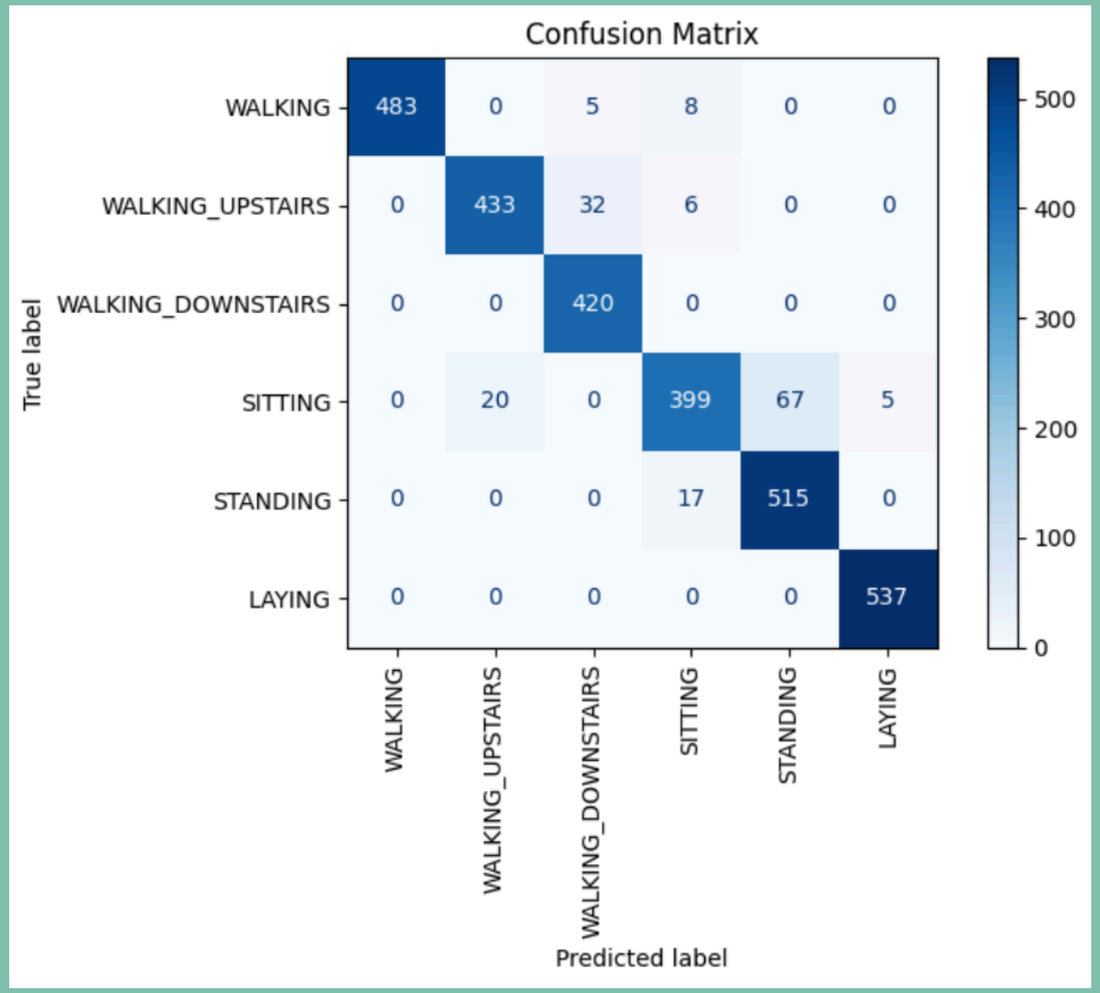


20 EPOCHS



500 EPOCHS

Hyperparameter Tuning 1D CNN



100 EPOCHS

Epochs: 100

reduced Learning Rate: 0.0005

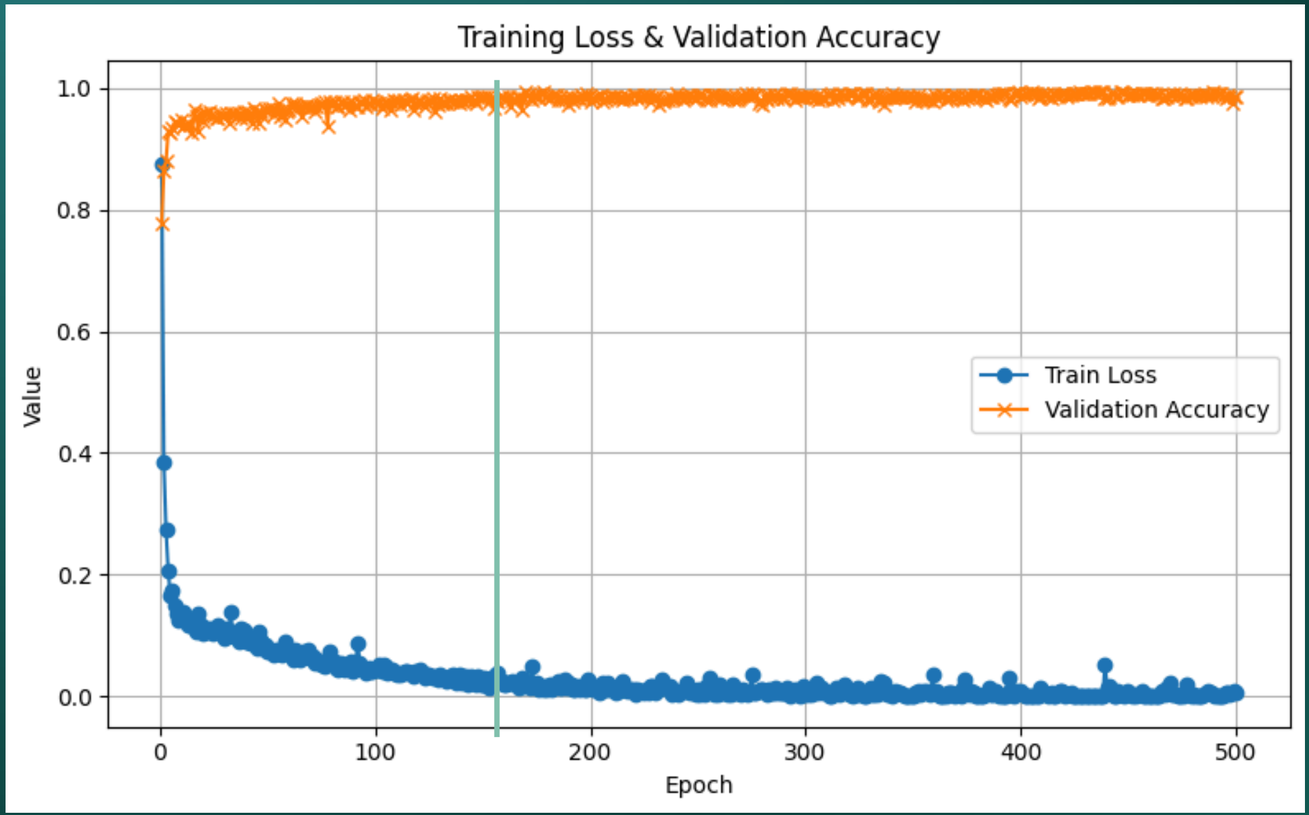
Loss Function: CrossEntropyLoss (with weighting)

- boosted Sitting and Standing

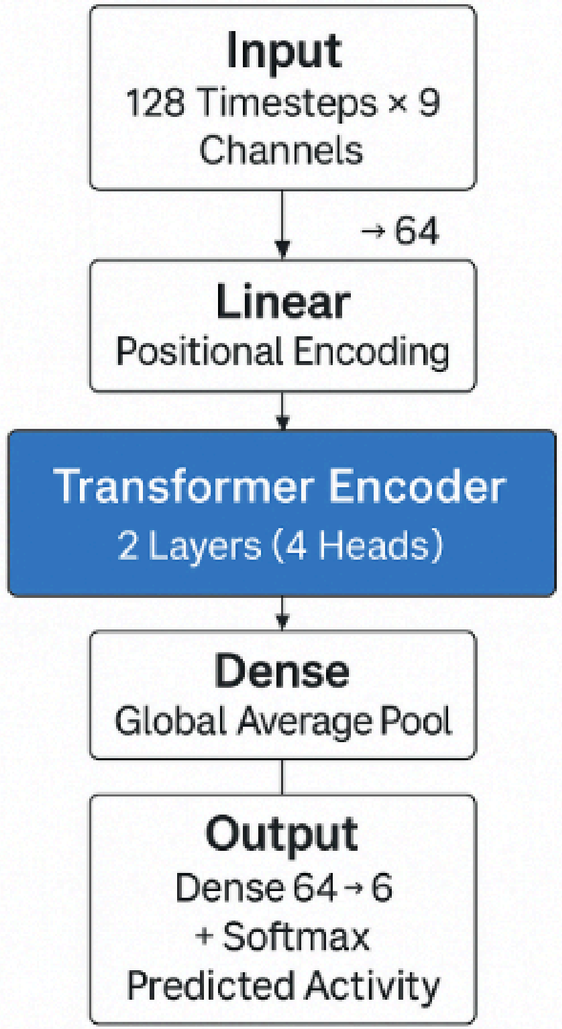
What it Learns

Full-sequence attention + global context

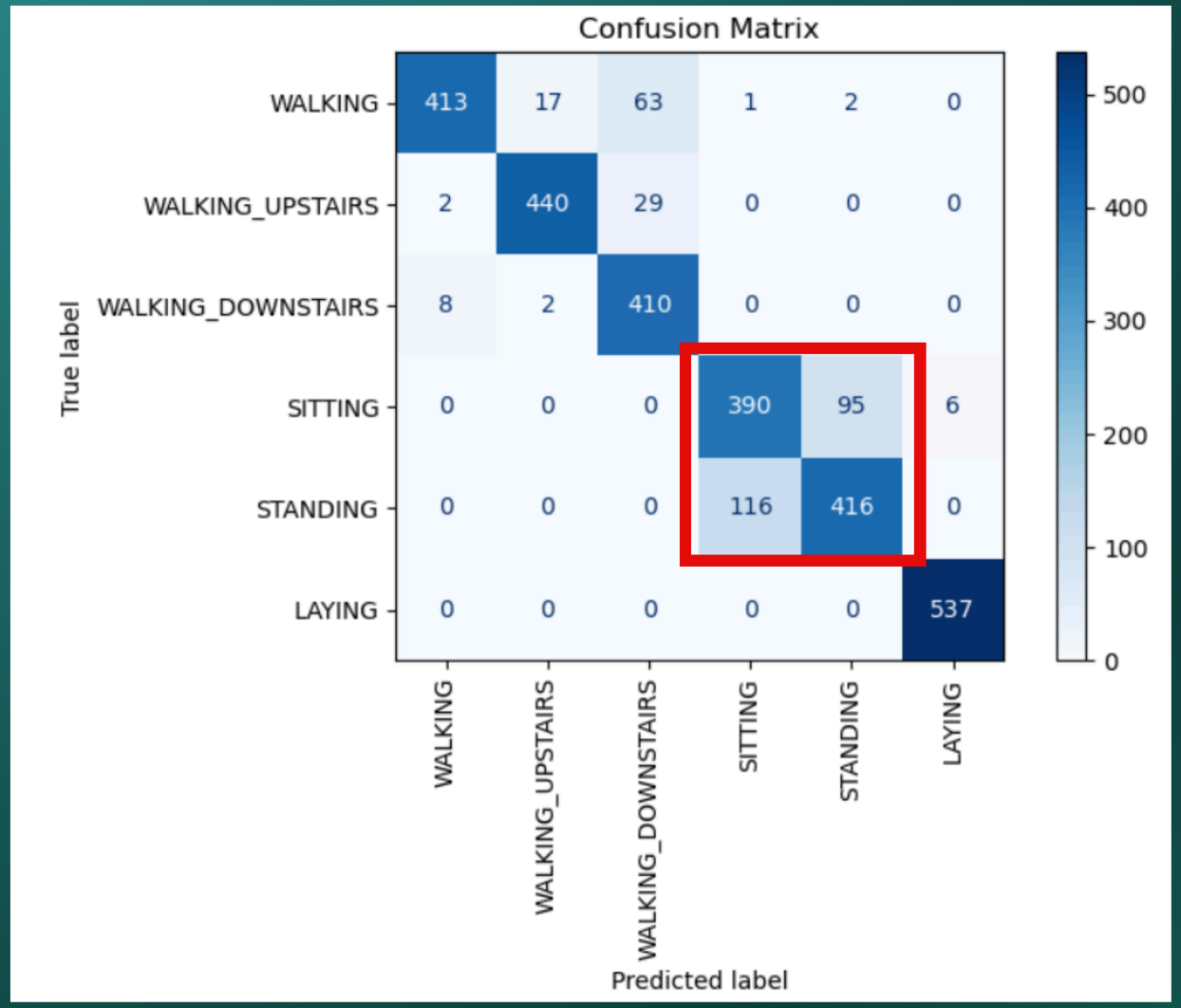
Best for long, complex motion patterns



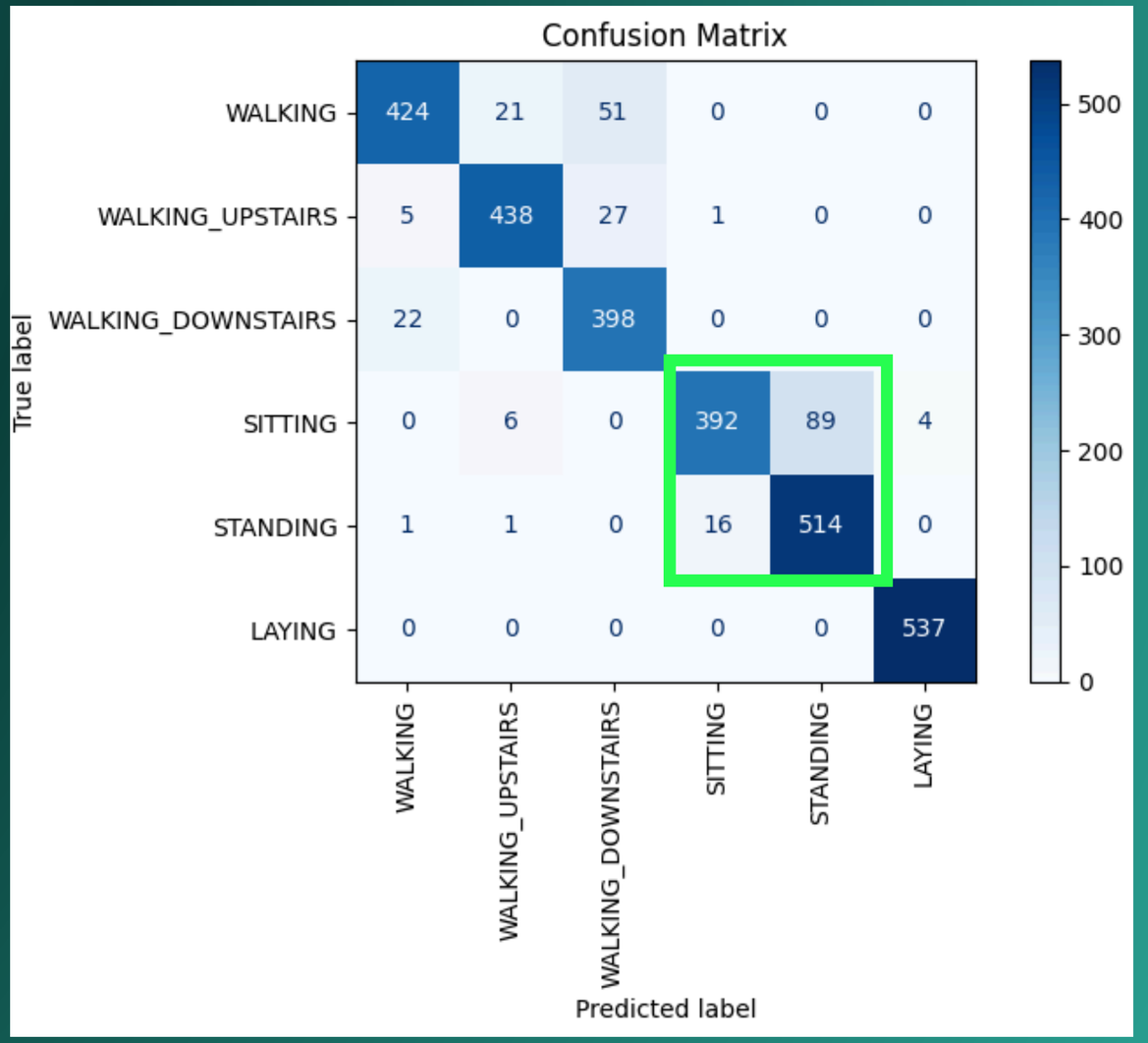
Transformer Encoder Model



Transformer



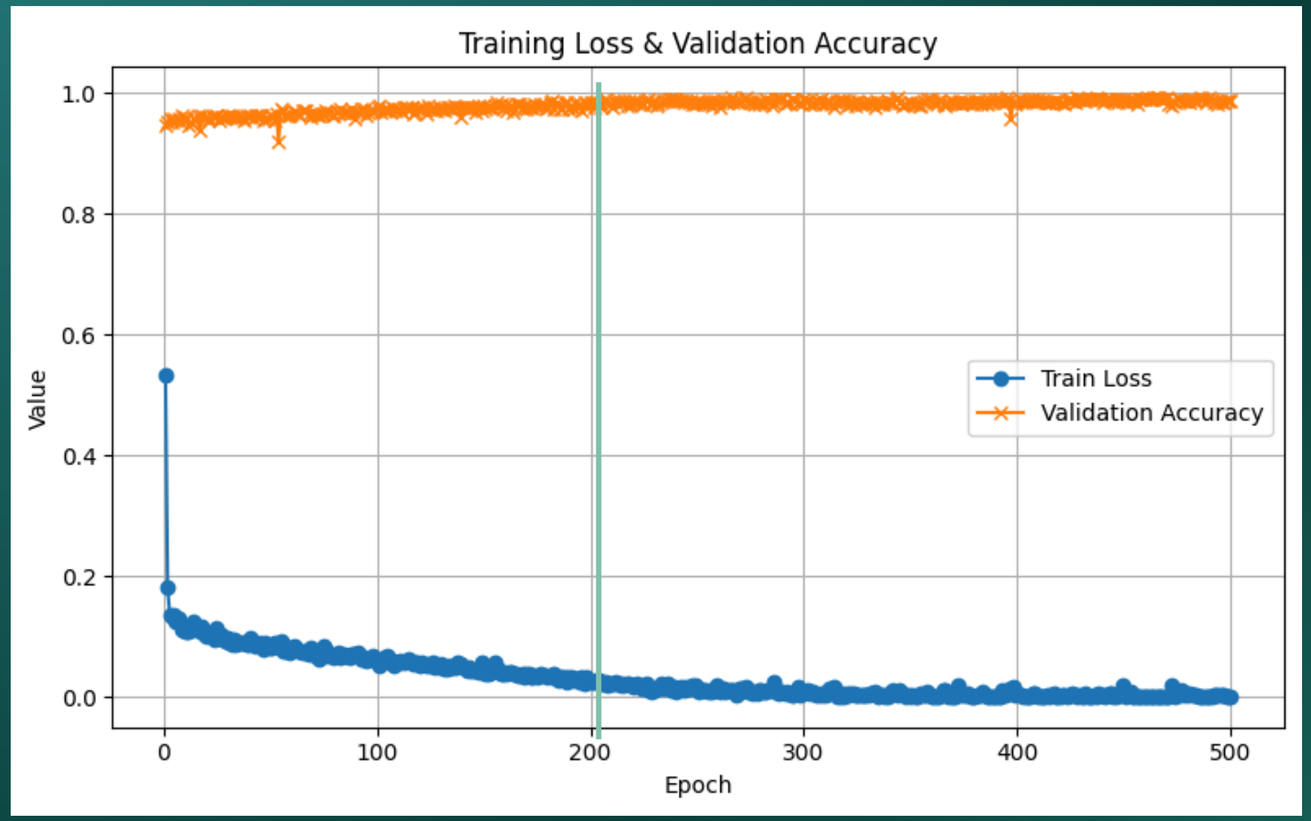
20 EPOCHS



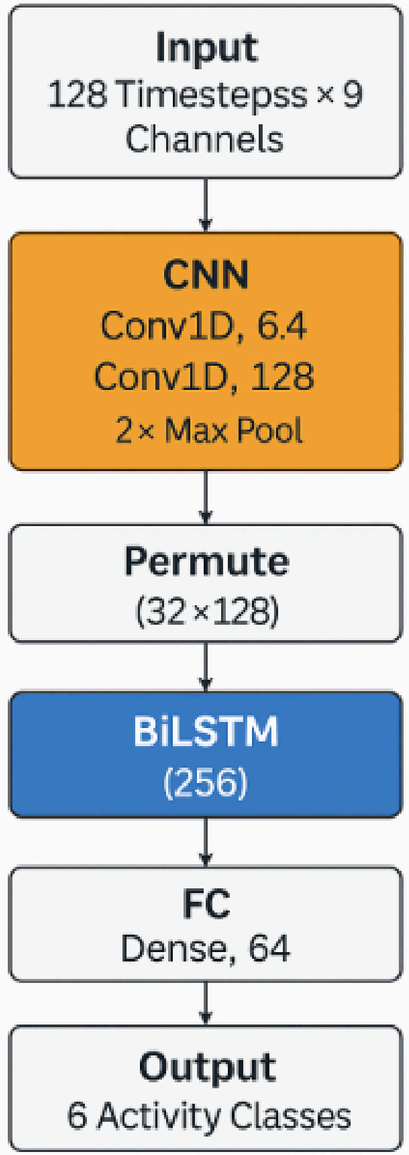
500 EPOCHS

What it Learns

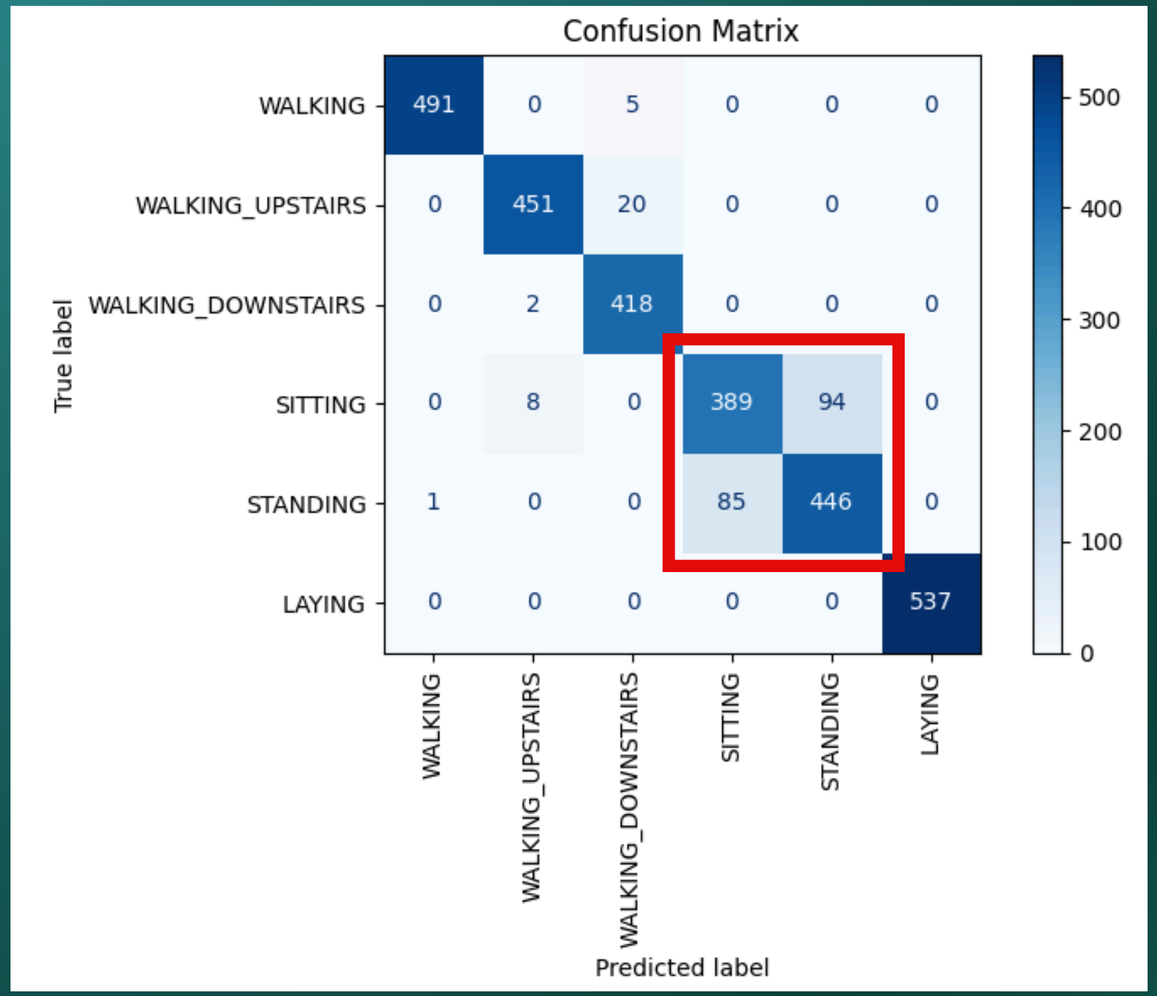
- Local patterns + temporal dependencies
- Balanced: local + sequence context
- Slower than pure CNN



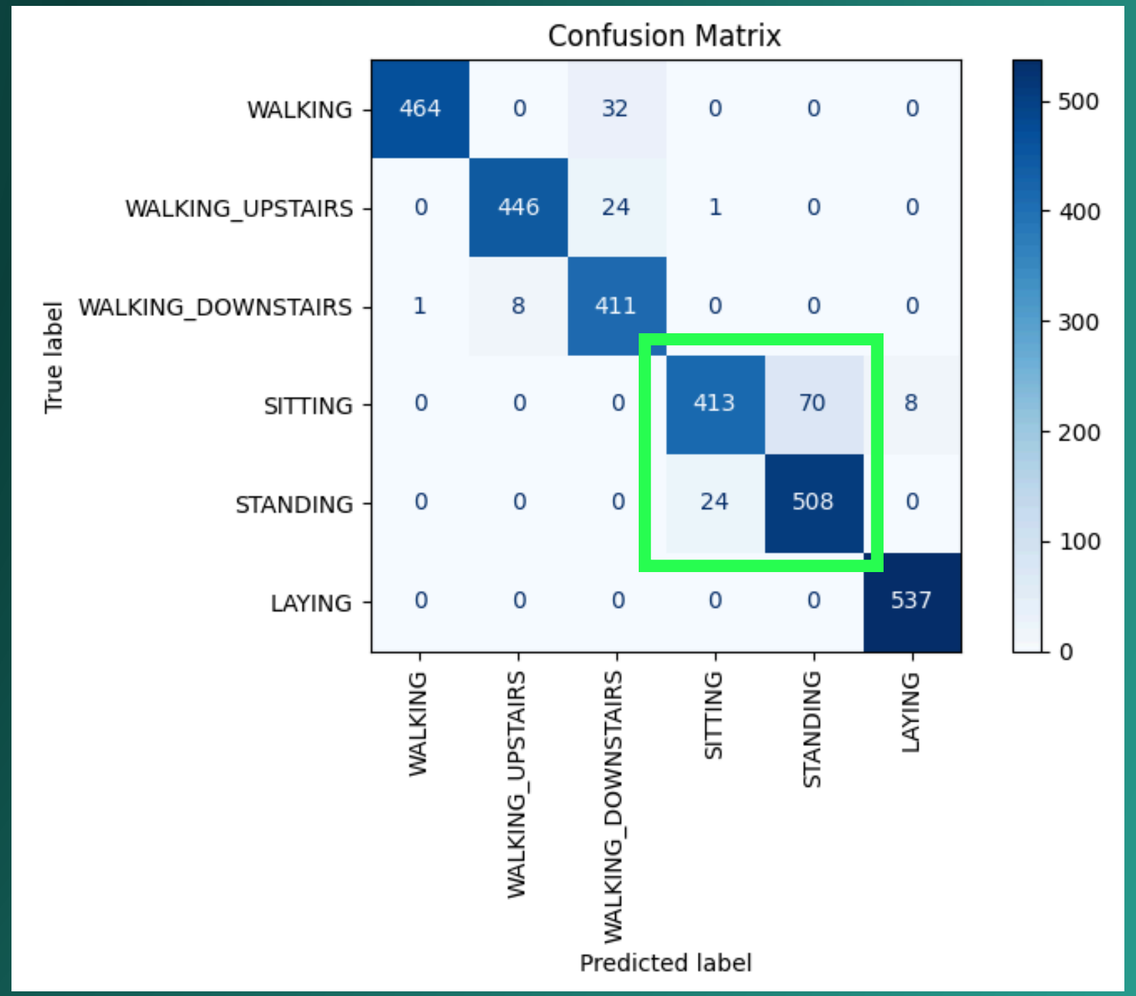
CNN-LSTM Model Architecture



CNN + LSTM Layer



20 EPOCHS



500 EPOCHS

Model	Accuracy	WALKING F1 Score	WALKING_ UPSTAIRS F1 Score	WALKING_ DOWNSTAIRS F1 Score	SITTING F1 Score	STANDING F1 Score	LAYING F1 Score
1D CNN	0.97	1	0.97	0.97	0.92	0.94	1
CNN-LSTM	0.94	0.97	0.96	0.93	0.89	0.92	0.99
Transformer	0.92	0.89	0.93	0.89	0.87	0.91	1

Key Takeaways

- 1D CNN gave the best performance overall
- Sitting vs. Standing was consistently the hardest to distinguish
- Hyperparameter tuning improved early training efficiency

Future Improvement

advanced data augmentation
early stopping
Add attention on top of CNN layers

Conclusion

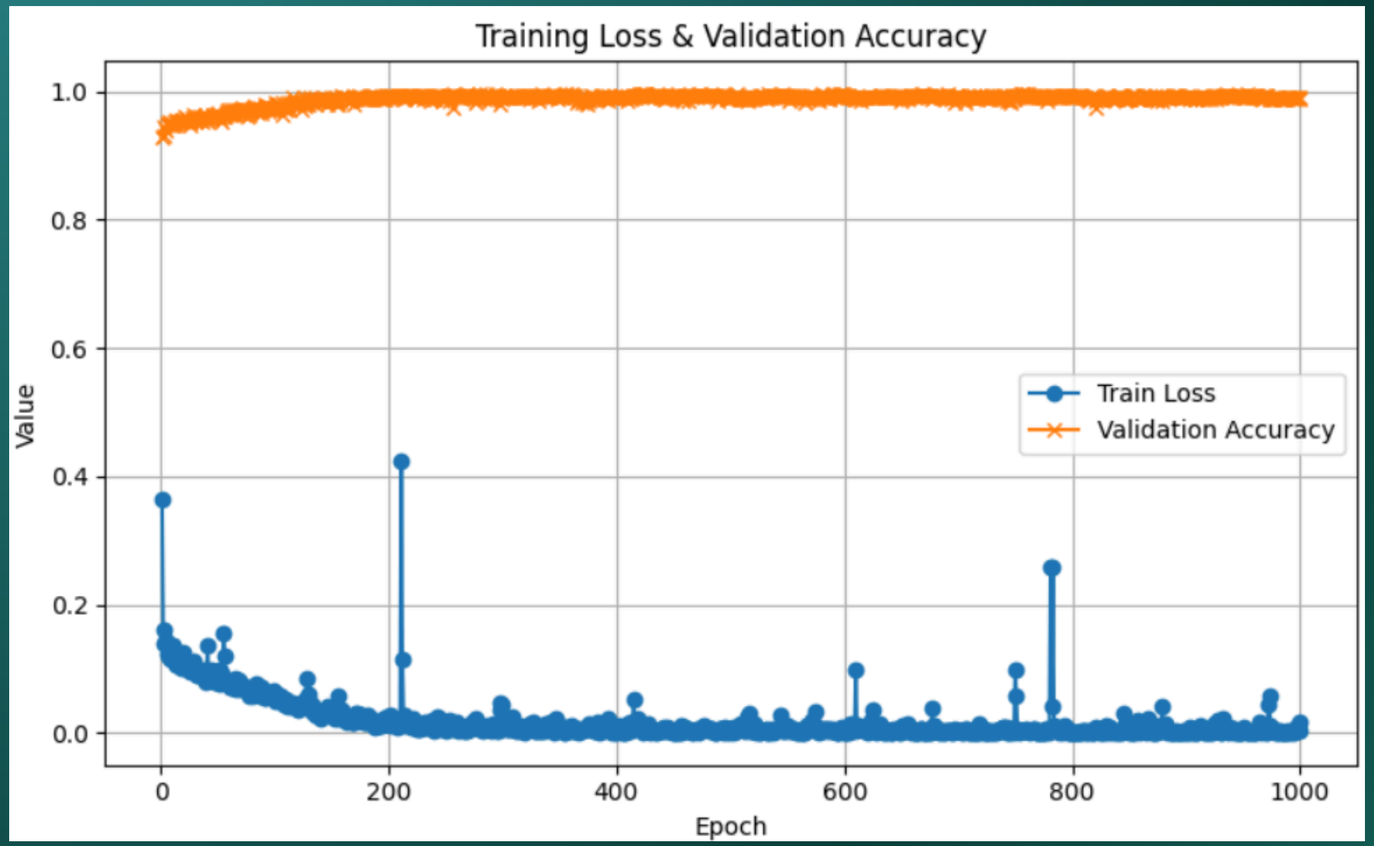
Questions?

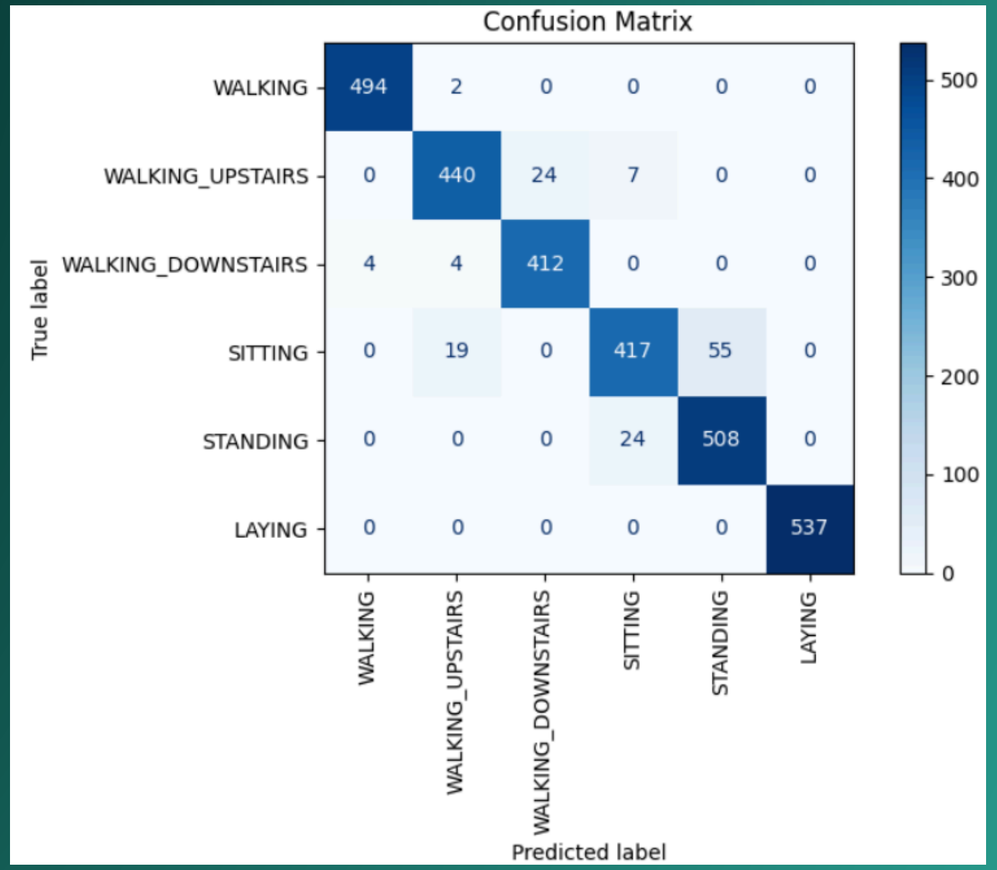
CNN (500 epochs)			
Activity	*Precision*	*Recall*	*F1 Score*
WALKING	1.00	1.00	1.00
WALKING_UPSTAIRS	0.97	0.97	0.97
WALKING_DOWNSTAIRS	0.97	0.98	0.97
SITTING	0.97	0.87	0.92
STANDING	0.91	0.98	0.94
LAYING	0.99	1.00	1.00
Accuracy			0.97

CNN + LSTM (500 epochs)			
Activity	*Precision*	*Recall*	*F1 Score*
WALKING	1.00	0.94	0.97
WALKING_UPSTAIRS	0.98	0.95	0.96
WALKING_DOWNSTAIRS	0.88	0.98	0.93
SITTING	0.94	0.84	0.89
STANDING	0.88	0.95	0.92
LAYING	0.99	1.00	0.99
Accuracy			0.94

Transformer (500 epochs)			
Activity	*Precision*	*Recall*	*F1 Score*
WALKING	0.94	0.85	0.89
WALKING_UPSTAIRS	0.94	0.93	0.93
WALKING_DOWNSTAIRS	0.84	0.95	0.89
SITTING	0.96	0.80	0.87
STANDING	0.85	0.97	0.91
LAYING	0.99	1.00	1.00
Accuracy			0.92

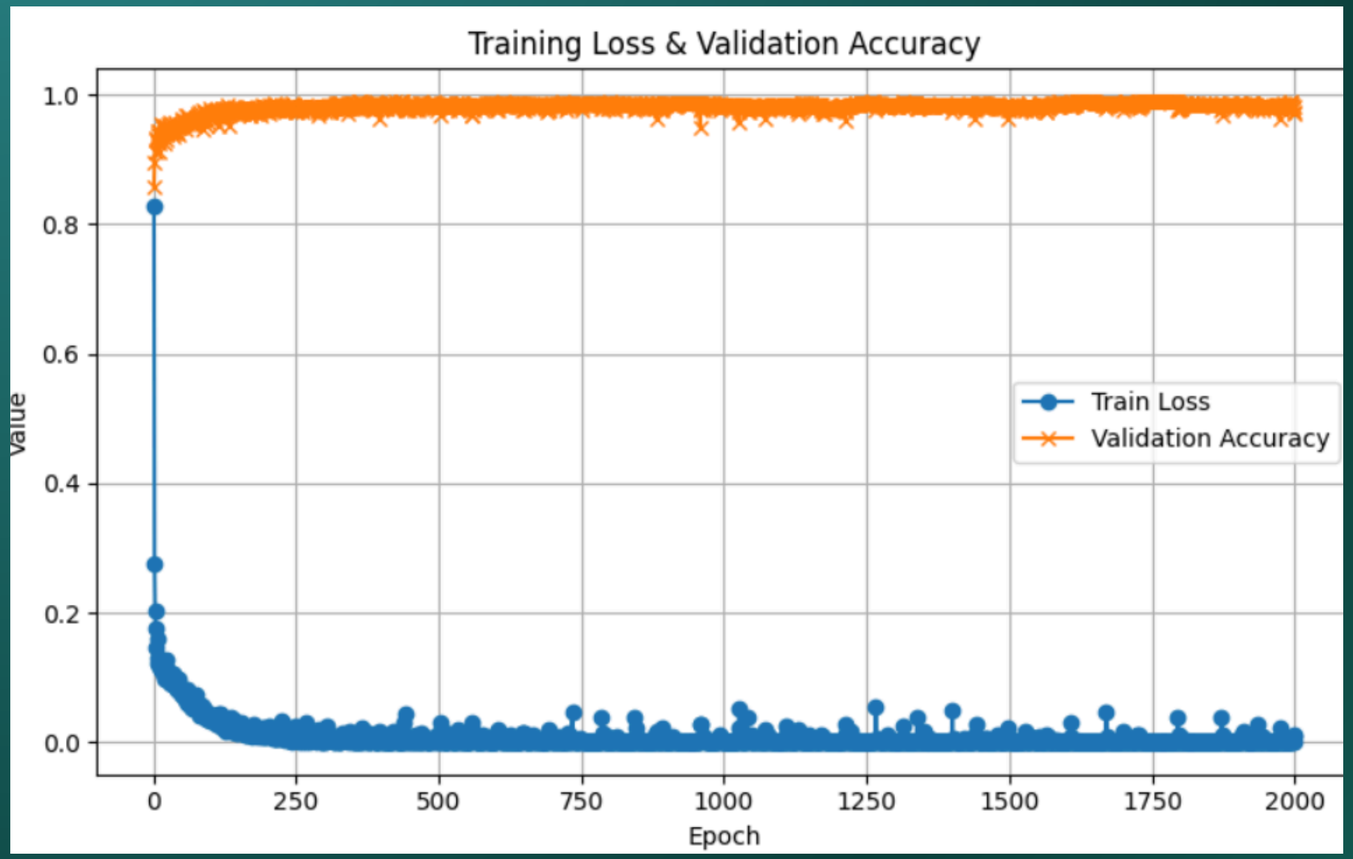
1D CNN

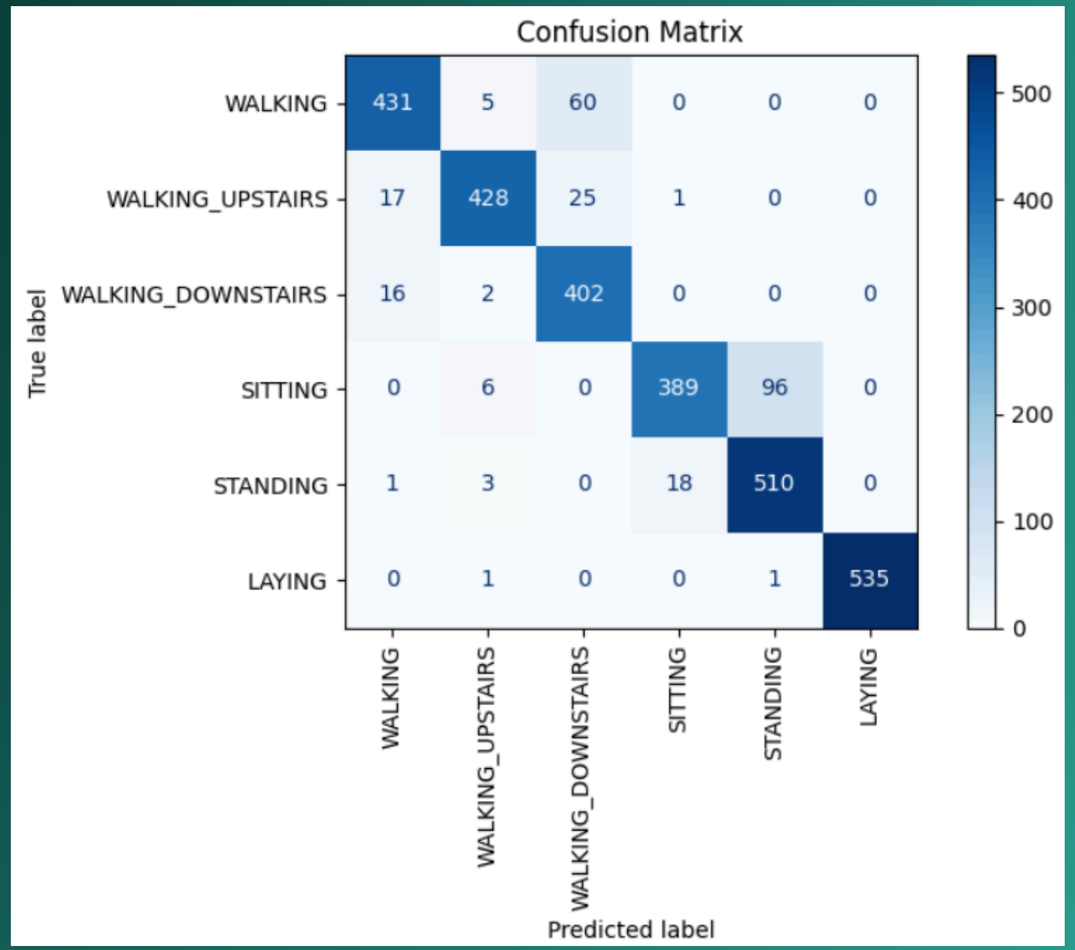




1000 EPOCHS

Transformer





2000 EPOCHS