



# A Convolutional Approach to Time Series Similarity

Divya Shanmugam, Davis Blalock, John Gutttag

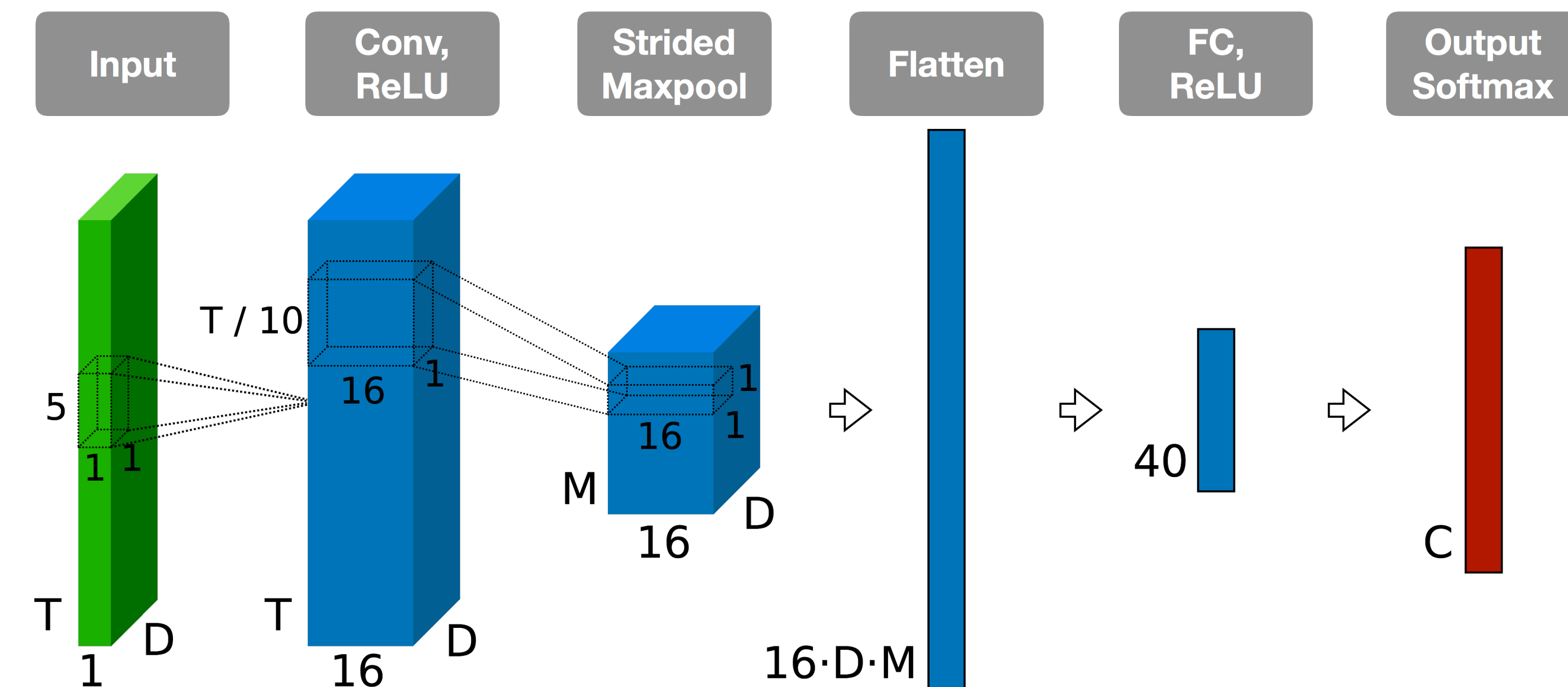
## Motivation

Scalable, accurate methods for measuring similarity between multivariate time series are valuable towards **patient similarity**, **disease classification**, and **anomaly detection**.

### Problem Statement

Given multivariate time series dataset  $D$ , produce an embedding where Euclidean distance accurately represents similarity between examples.

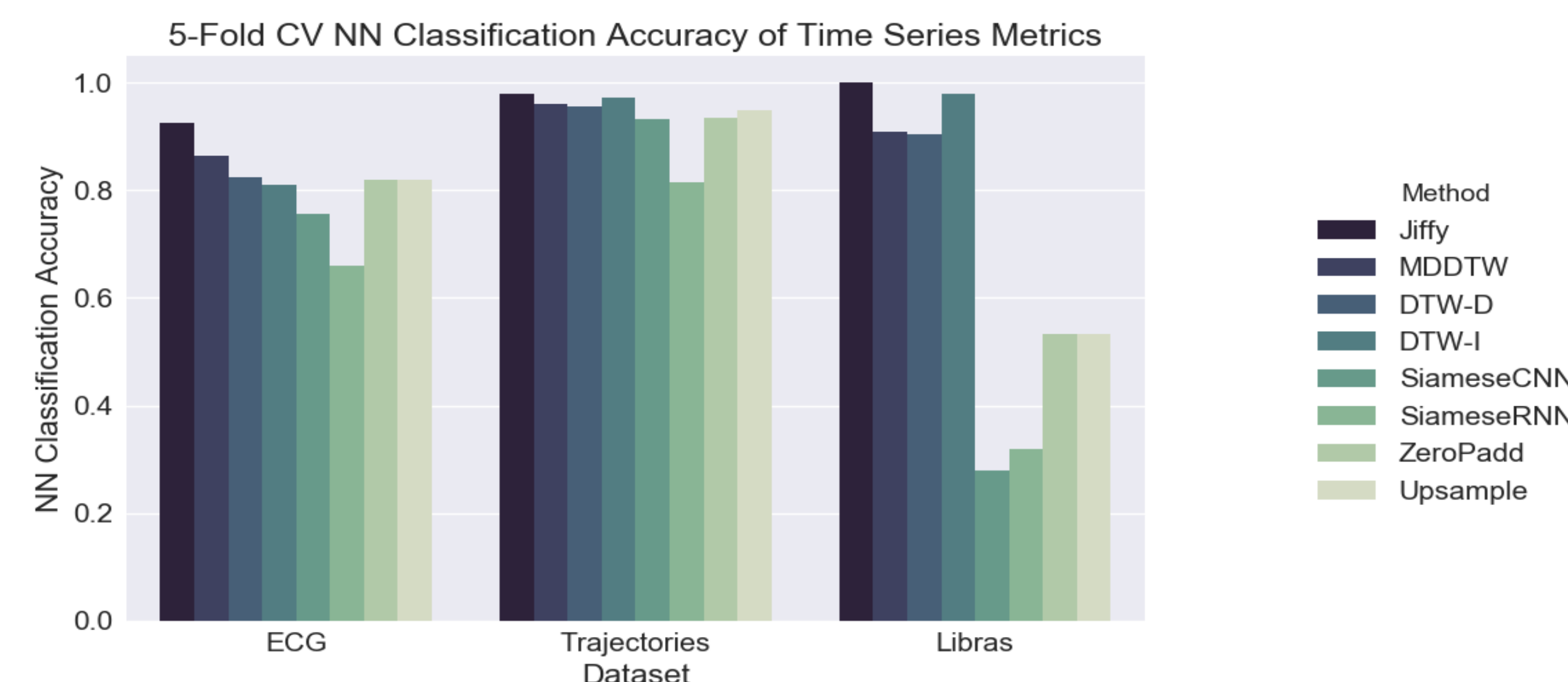
## Architecture



- Trained with cross-entropy loss

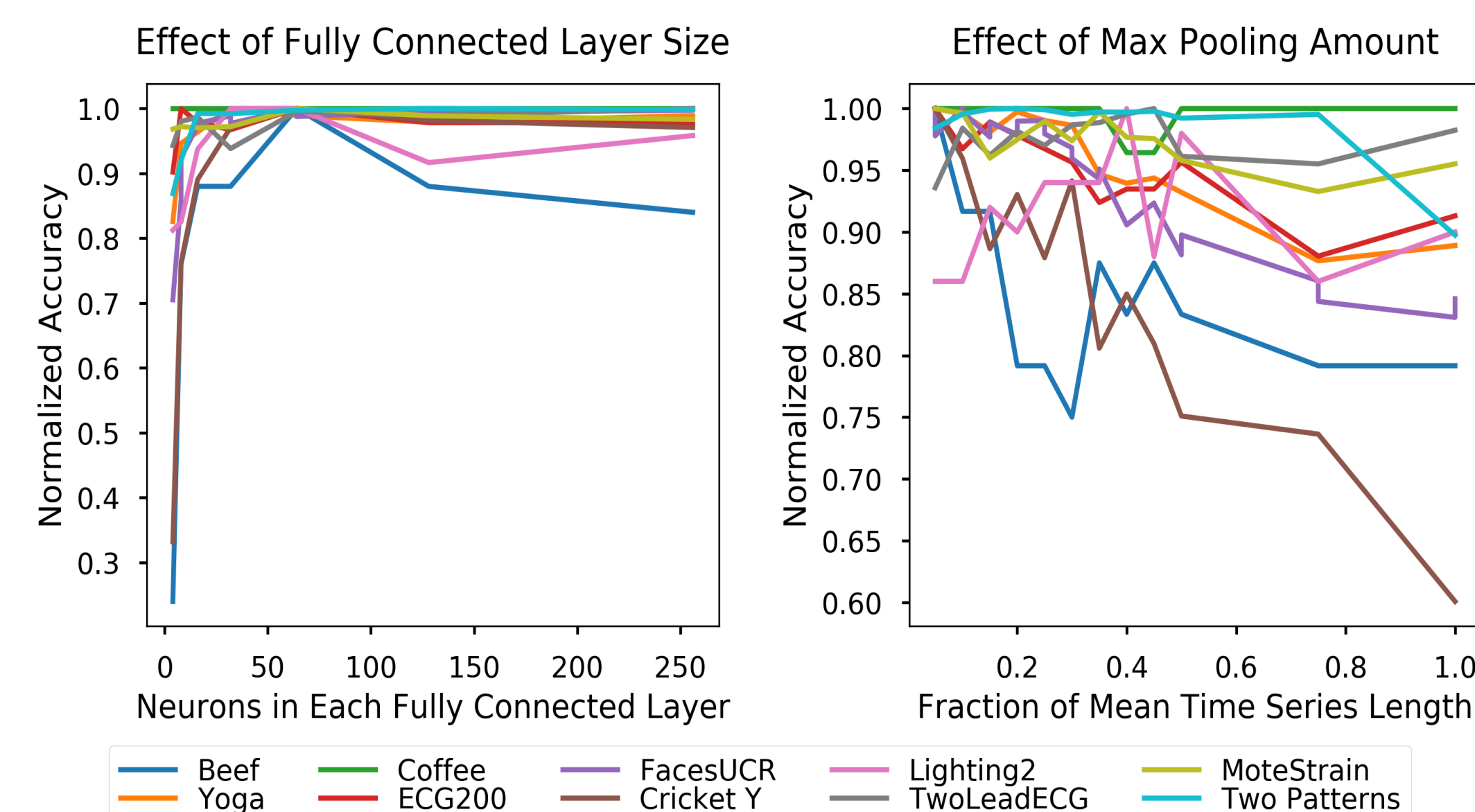
## Results

- Jiffy equals or exceeds accuracies of all competing methods



## Parameter Stability

- No benefit to larger embedding
- Excels with 20-40% maxpooling



- Tested across UCR Archive datasets

## Clustering [WIP]

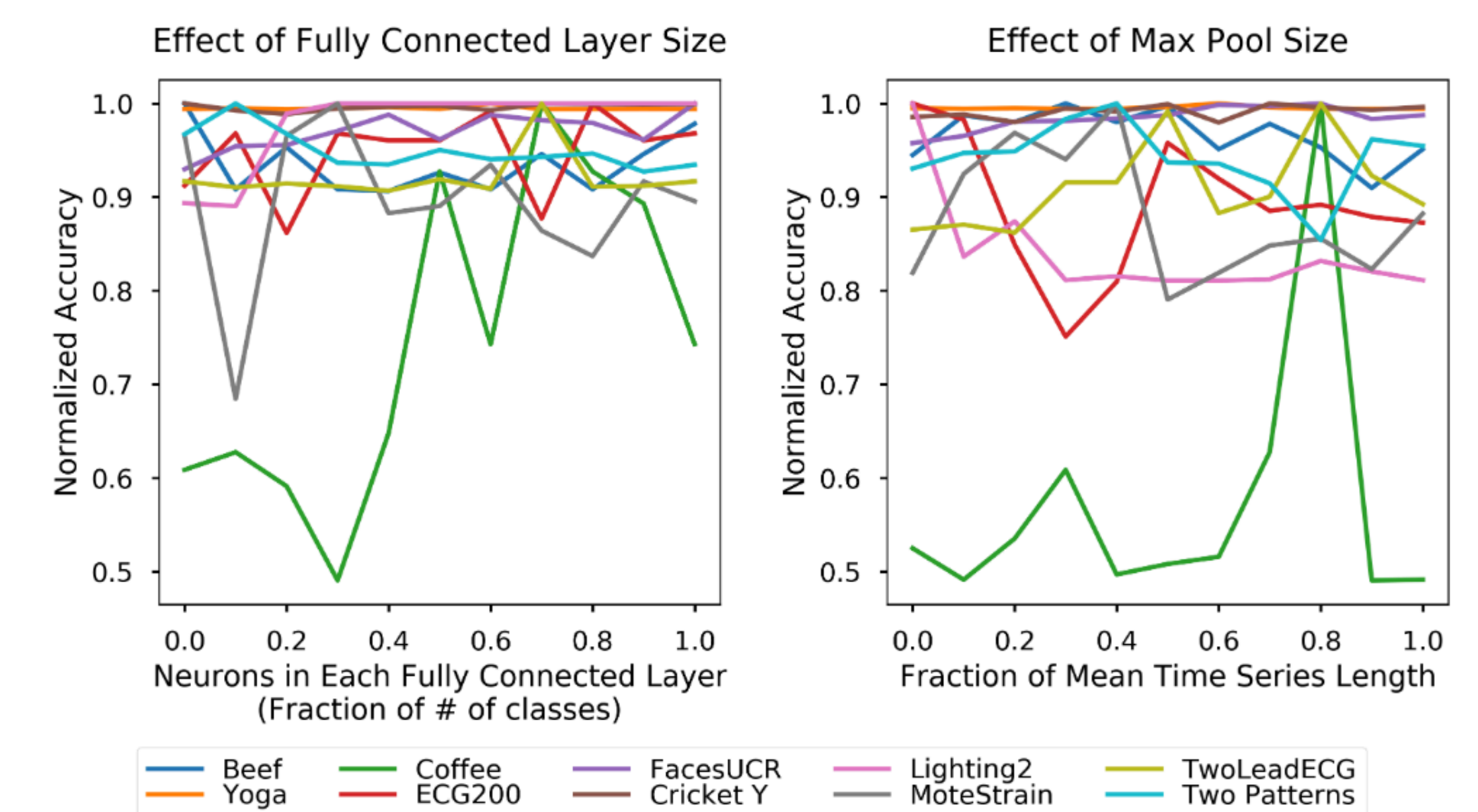
- Problem: Given a set of unlabeled multivariate time series, learn a latent representation that demonstrates high clustering purity
- Approach: Apply convolutional autoencoder to discover patterns important to a latent representation

### Preliminary Results

Dataset	Jiffy	LDPS	SPIRAL	Zero-Padding	Upsampling
ArabicDigits	0.849	0.826	0.870	0.887	<b>0.928</b>
AUSLAN	<b>0.990</b>	<b>0.990</b>	0.868	0.982	0.982
CharacterTrajectories	<b>0.965</b>	0.939	0.886	<b>0.965</b>	0.959
ECG	0.586	0.550	0.505	0.581	<b>0.627</b>
Libras	0.883	0.870	0.881	<b>0.885</b>	<b>0.885</b>
Wafer	<b>0.746</b>	0.554	0.448	0.500	0.596
<b>Mean Rank</b>	<b>2.0</b>	3.67	4.5	2.5	<b>1.83</b>

### Preliminary Parameter Stability

- Small FC layer size is sufficient
- Aggressive maxpooling maintains accuracy



## Current work

- Qualitatively analyze strengths/weaknesses of convolutional autoencoder time series representation
- Operate MV time series embedding approach on ECG signals

## Solution

We present Jiffy, a method capable of producing a data dependent embedding for multivariate time series.

- Task reframed as representation learning
- Through aggressive maxpooling + downsampling, network remains compact and accurate

## Datasets

- 6 multivariate time series datasets derived from sign language motion, pen movement, heartbeats, and speech

Table 1: Summary of Multivariate Time Series Datasets.

Dataset	# Variables	# Classes	Length	# Time Series
Libras	2	15	45	360
AUSLAN	22	25	47-95	675
CharacterTrajectories	3	20	109-205	2858
ArabicDigits	13	10	4 - 93	8800
ECG	2	2	39 - 152	200
Wafer	6	2	104 - 198	1194