



WROCŁAW UNIVERSITY
OF ENVIRONMENTAL
AND LIFE SCIENCES

Evaluating Dimensionality Reduction Techniques and Neural Network Architectures for High-Dimensional, Small-Sample Microbiome Classification

Marek Sztuka, Joanna Szyda

Goals

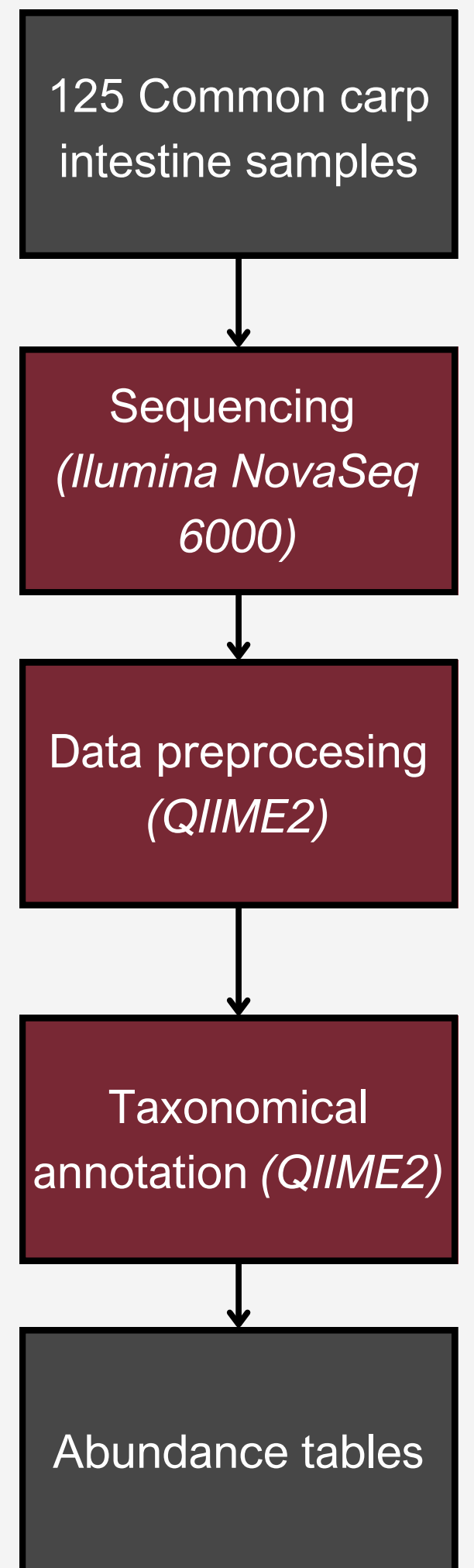
Deep Learning (DL) approaches in classification of high dimensional - small sample, sparse, biological data

Development of DL model capable of classification based on differences in effective microorganism supplementation

Materials

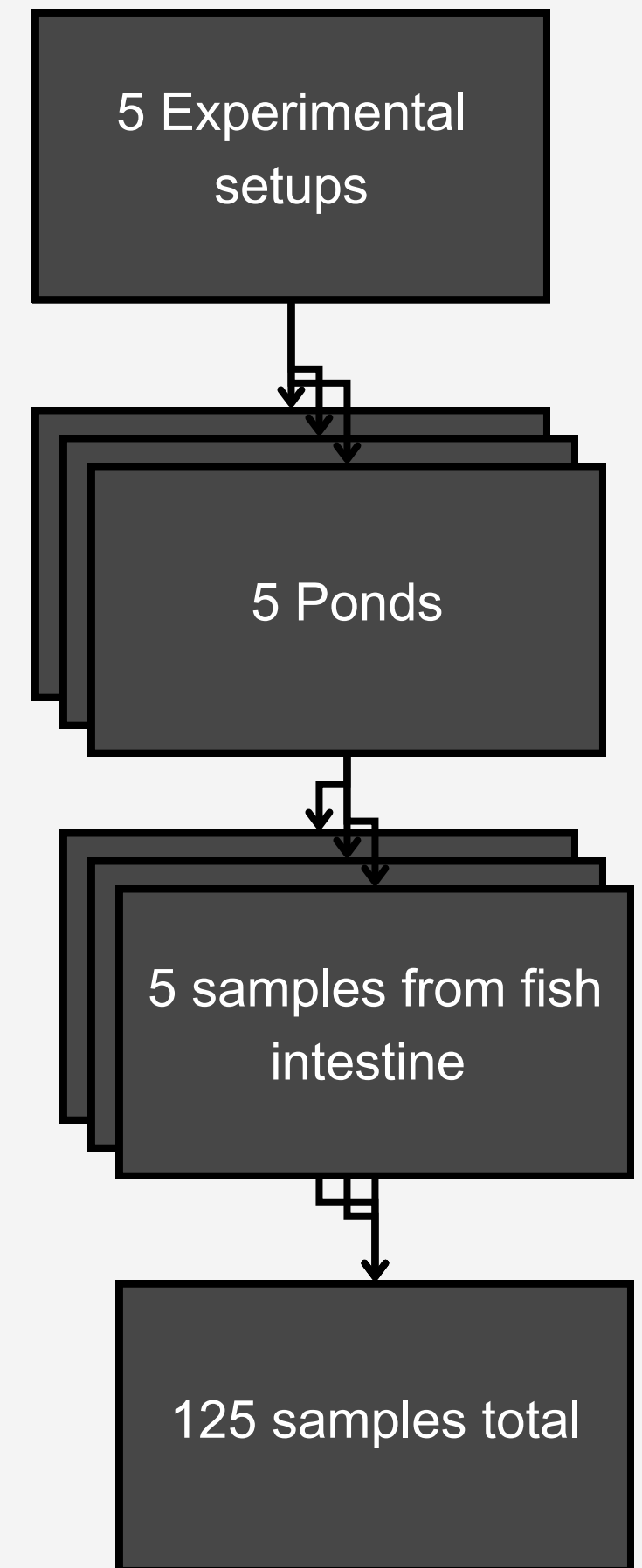
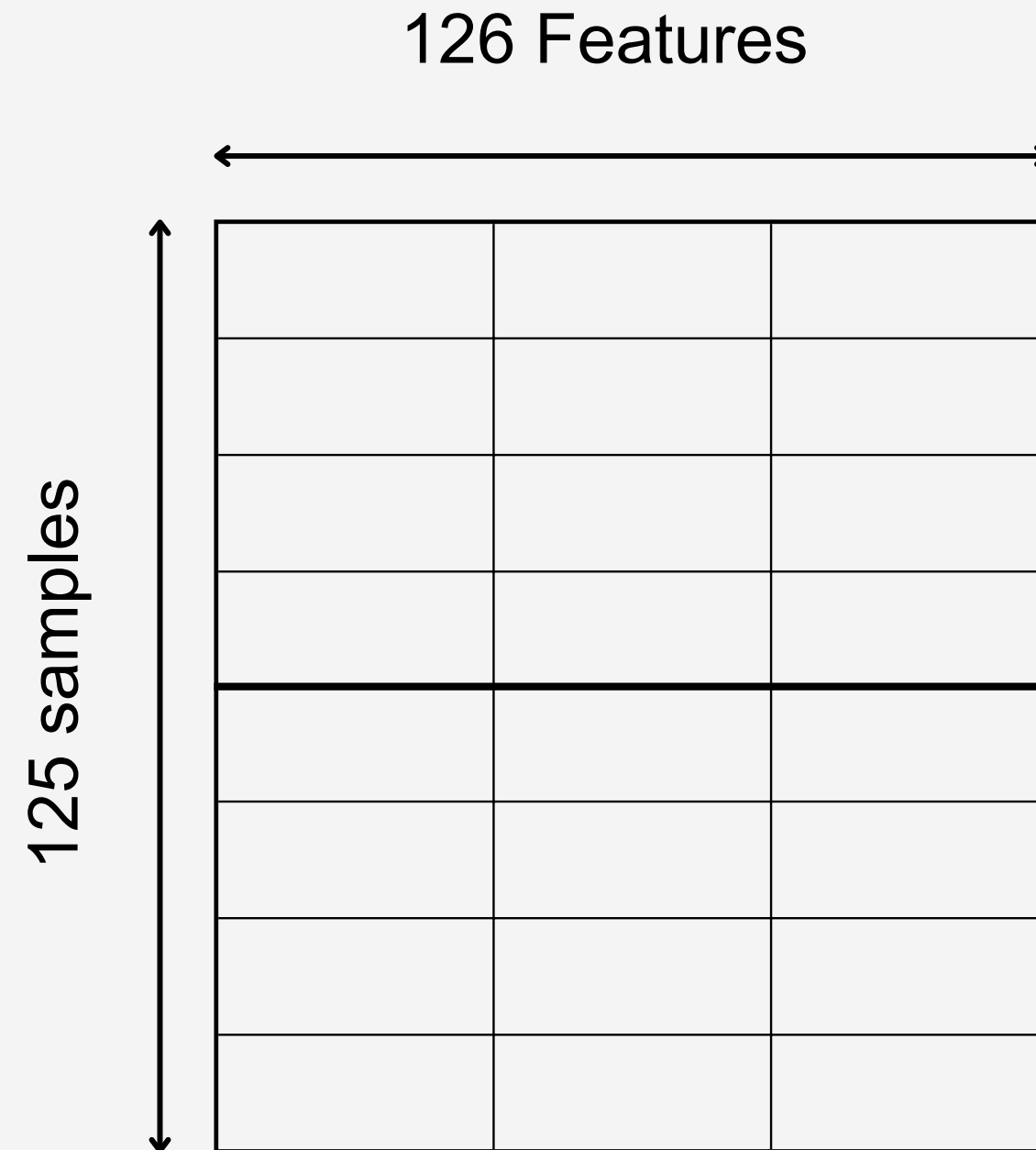
- Microbiome of fish intestine
- Sequenced reads of hypervariable regions of 16S rRNA gene
- 5 experimental setups, with different effective microorganism supplementation

Setup:	Control	1	2	3	4
Supplementation	X	X	F1	X	F2
Feed	X				
Enviromental	X	W1	W1	W2	W2

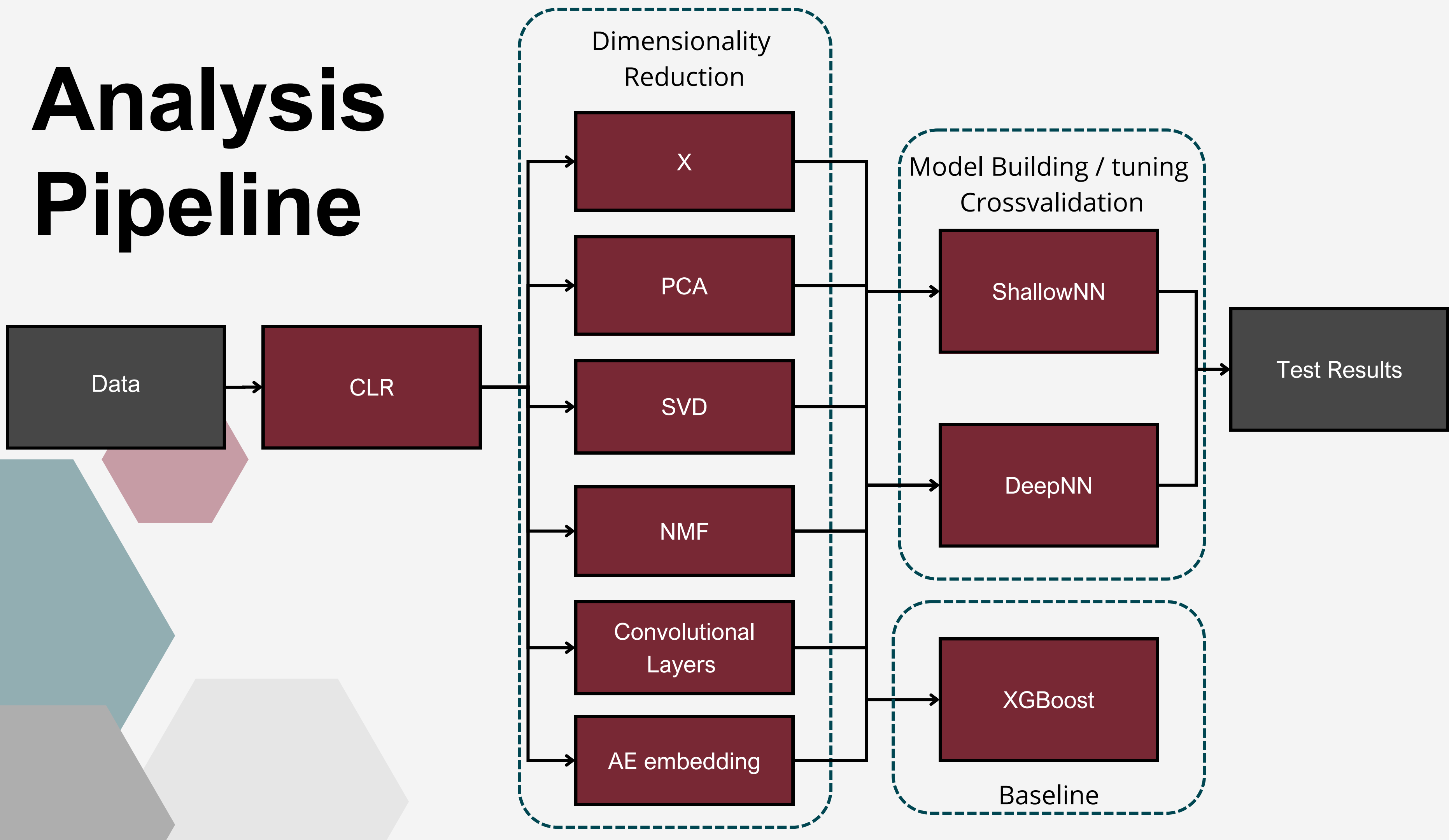


Materials

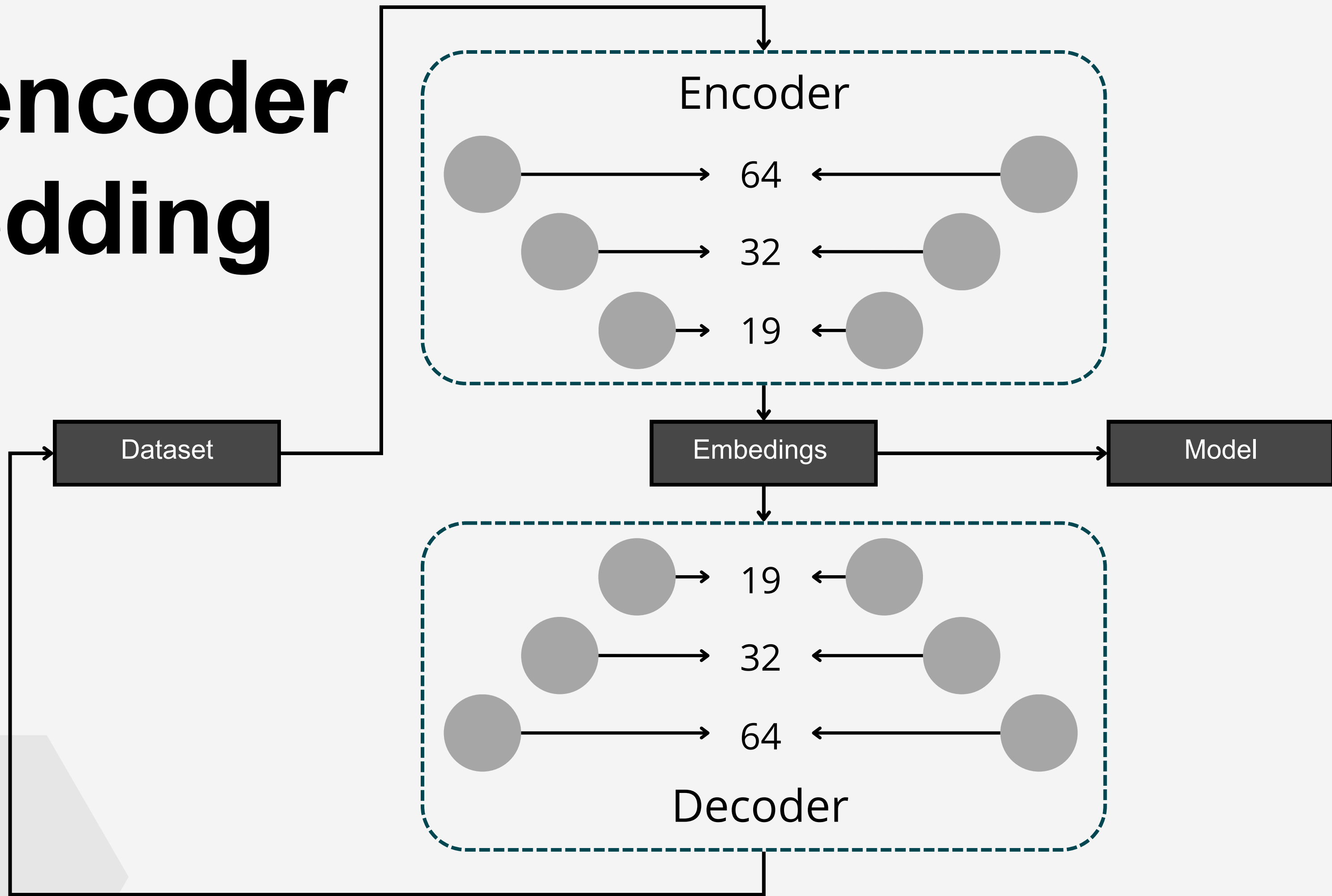
- Features - Bacteria families
- Samples - Individuals
- Cells - Bacteria abundance
- Sparse / zero inflated data



Analysis Pipeline

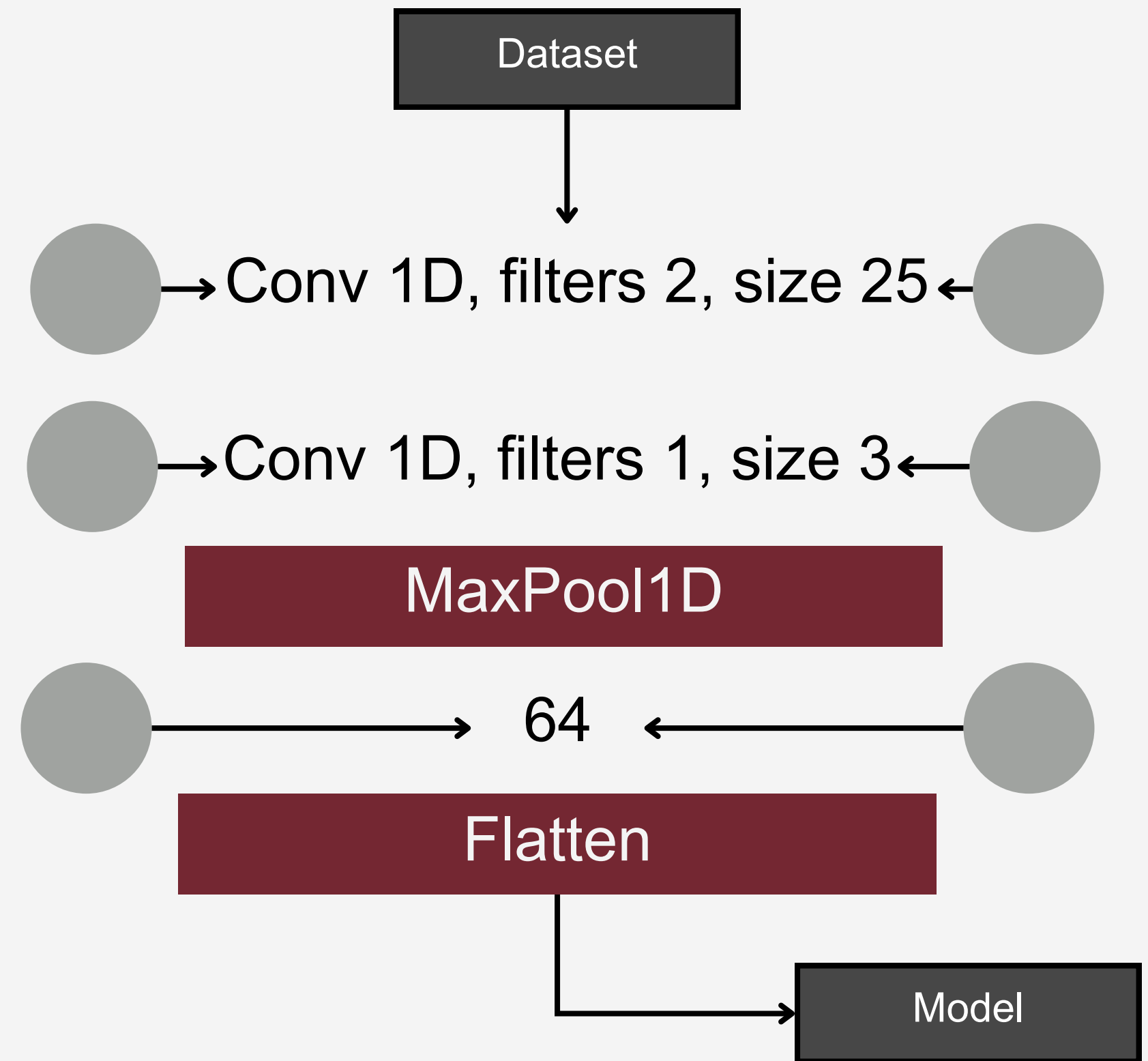


Autoencoder Embedding



Convolution Layers

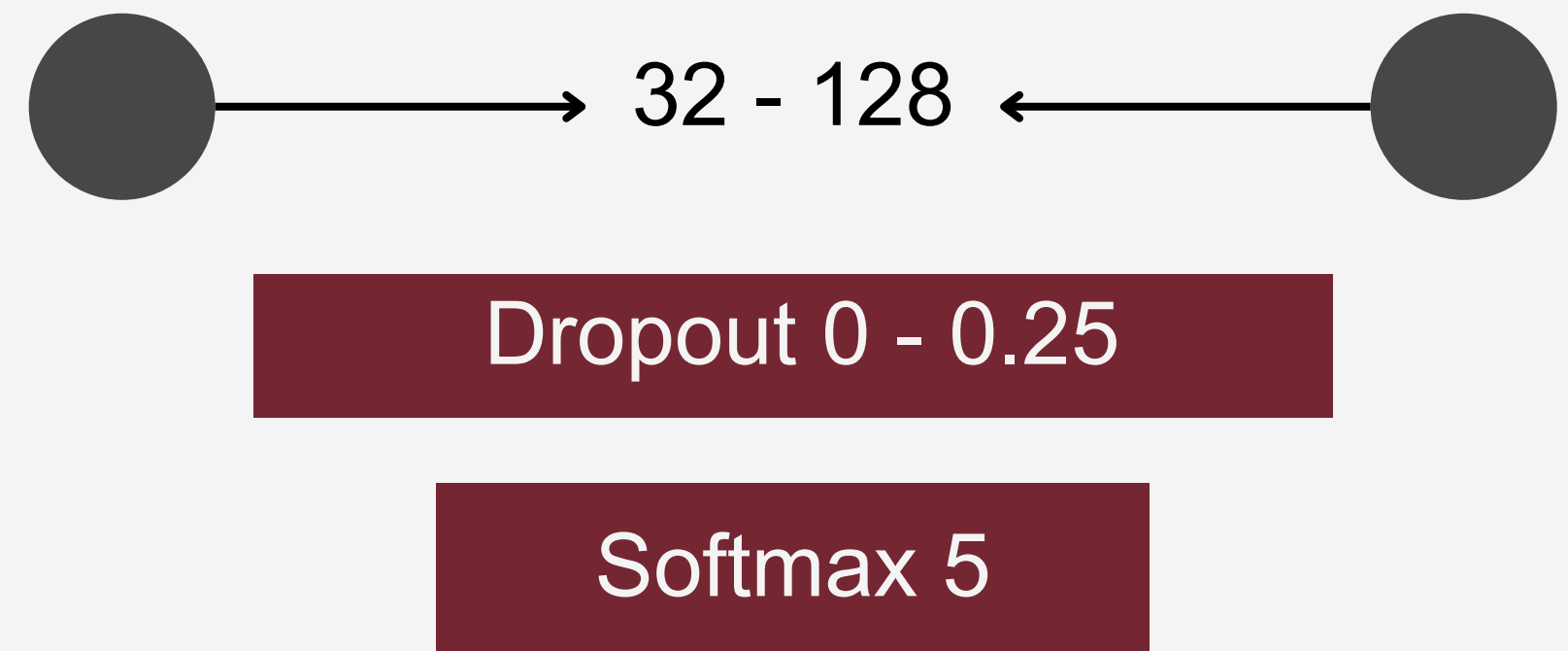
Fish - 1D Picture



Model Architecture

- ShallowNN

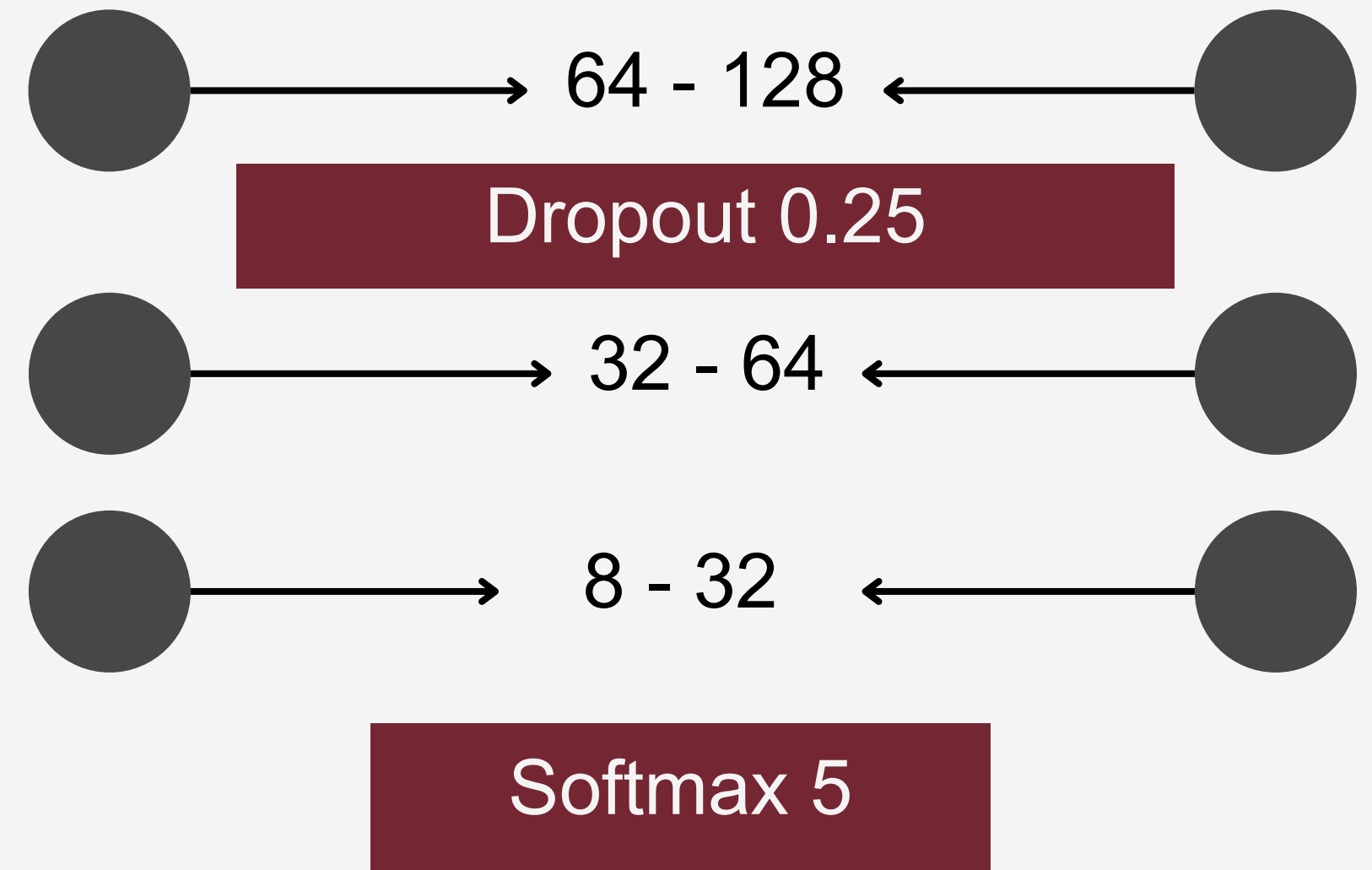
- Activation - ReLU
- Optimizer - ADAM
- Loss - Categorical Crossentropy
- Metric - Accuracy
- Weight regularizer - L2 (0.01)



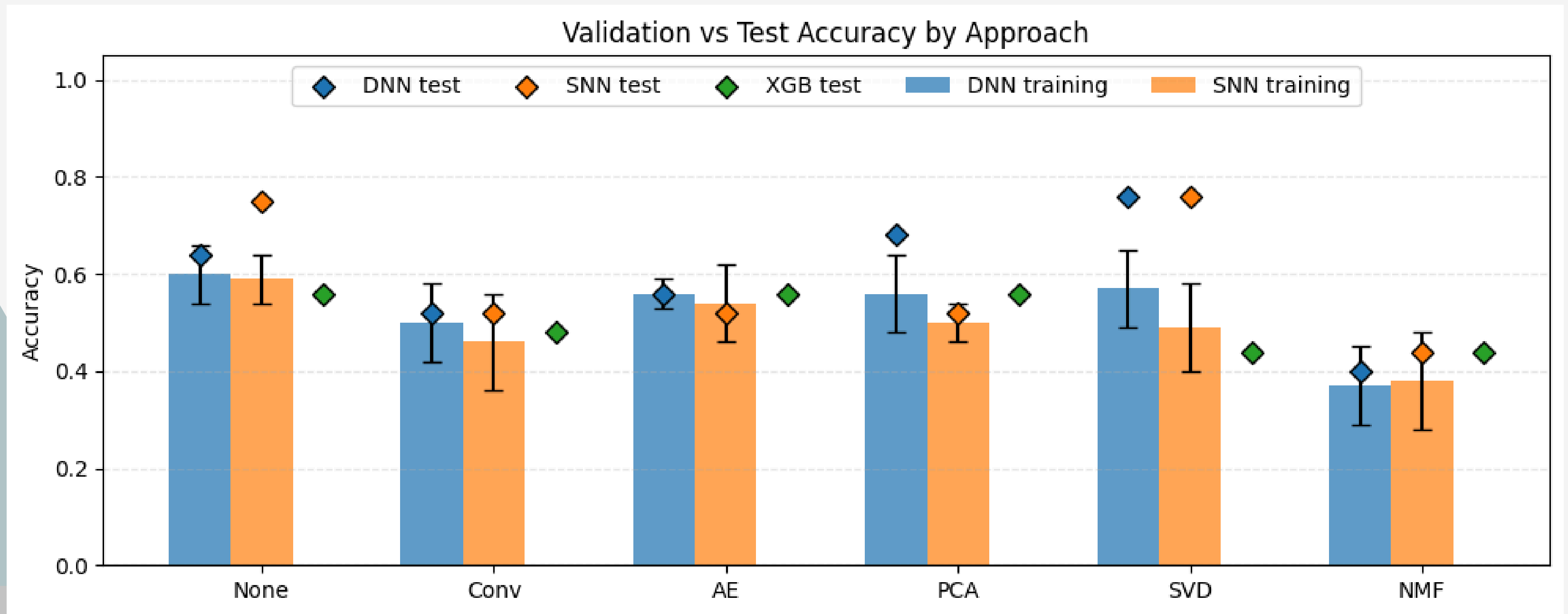
Model Architecture

- *DeepNN

- Activation - ReLU
- Optimizer - ADAM
- Loss - Categorical Crossentropy
- Metric - Accuracy
- Weight regularizer - L2 (0.01)



Results



Simulation

- Based on CLR transformed data
- Same shape
- Normal Distribution
- S - mean shift
- P_s - columns with shifted mean

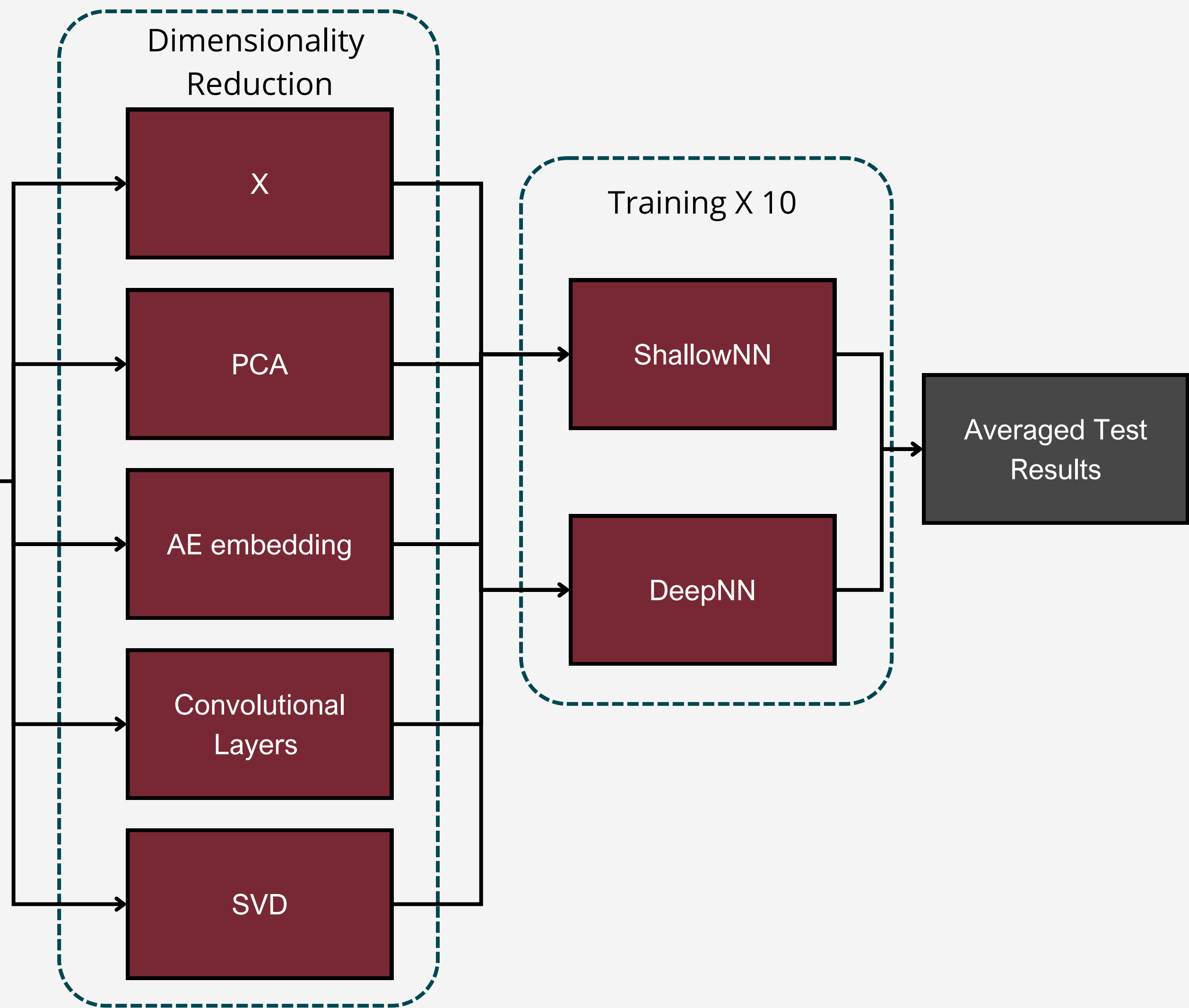
$$5 \xleftarrow{P_s} 125$$

$$0.1 \xleftarrow{S} 2.0$$

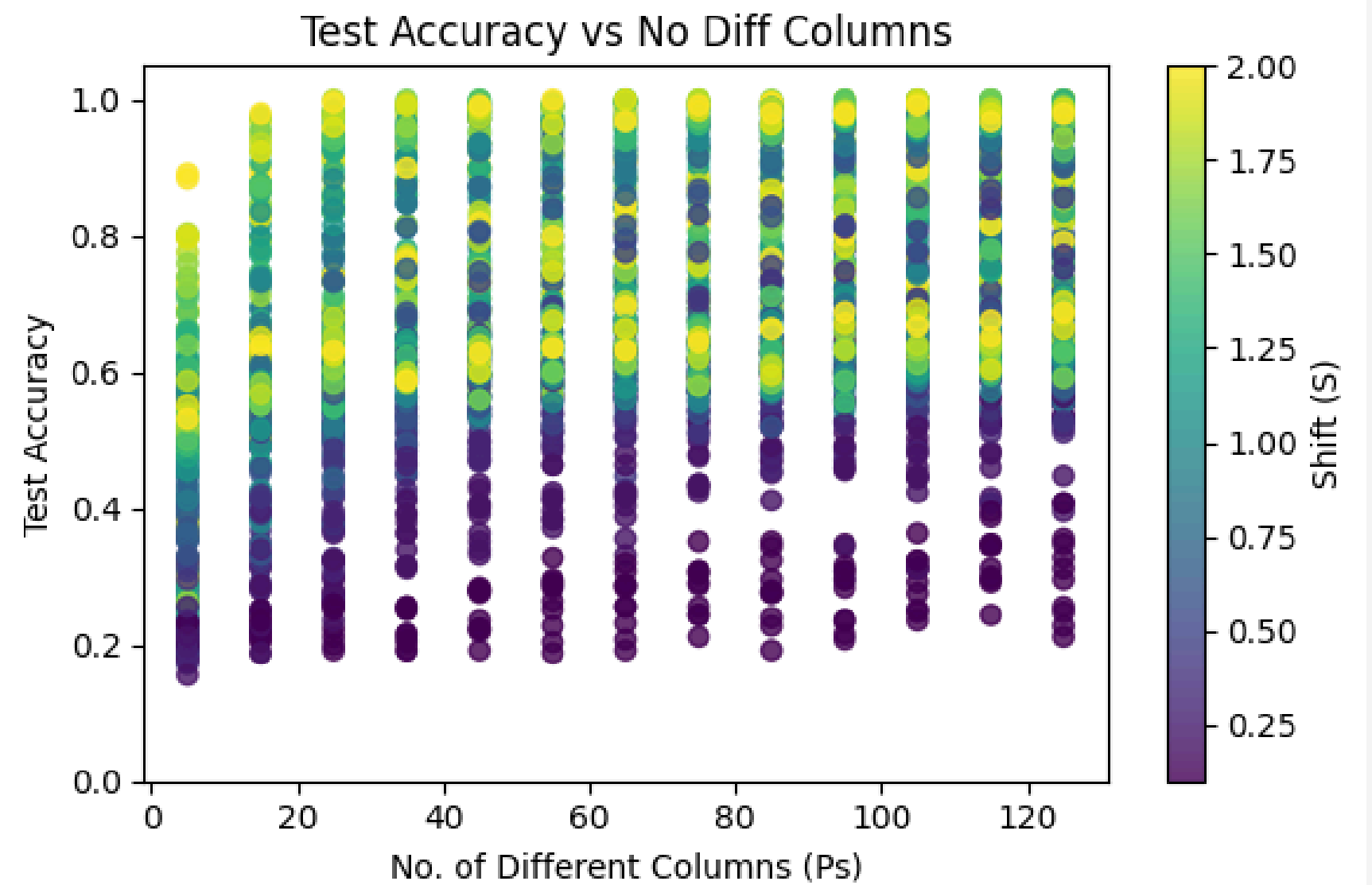
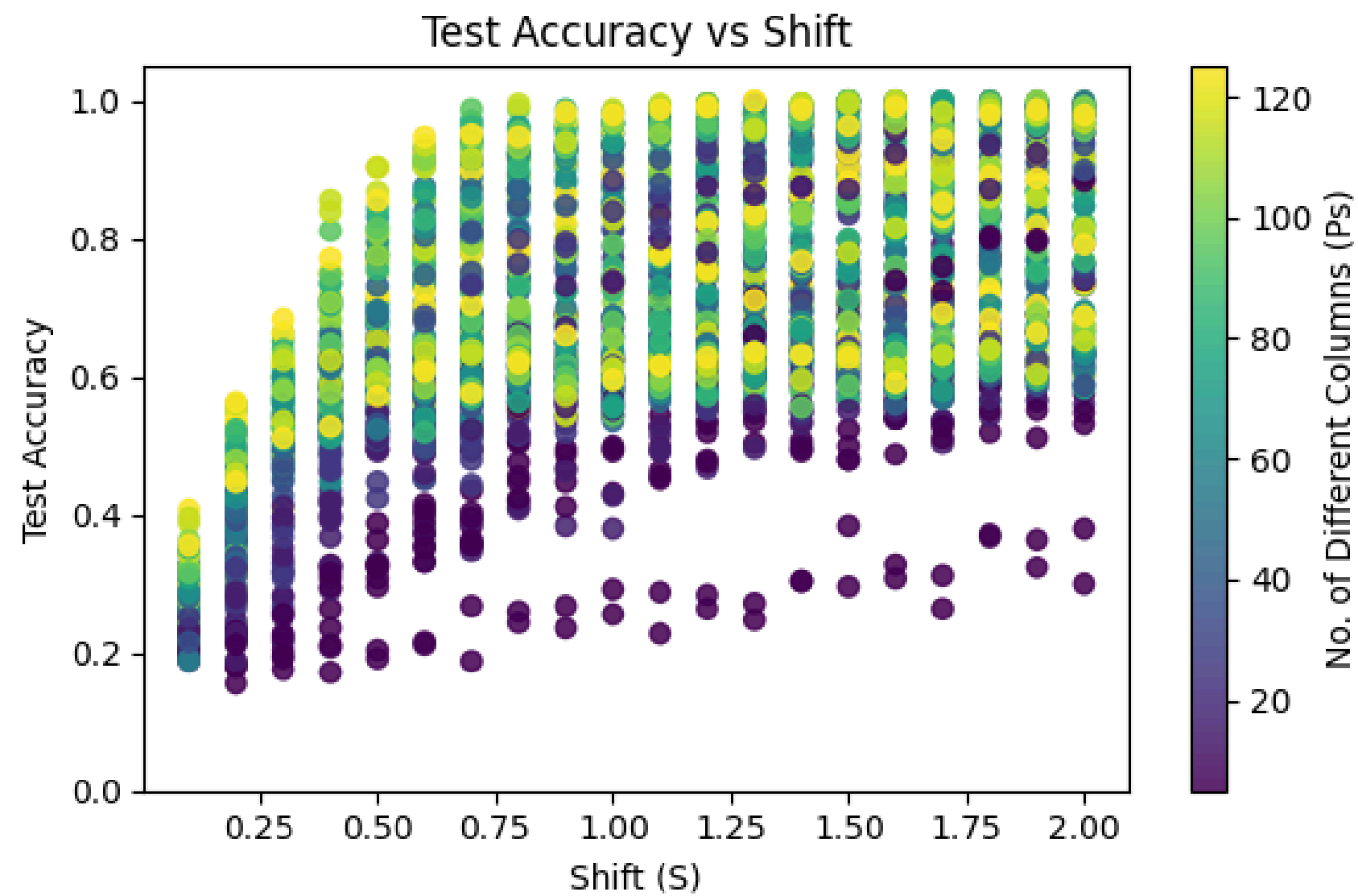
$\times 10$

Simulaiton Pipeline

Simulated Data

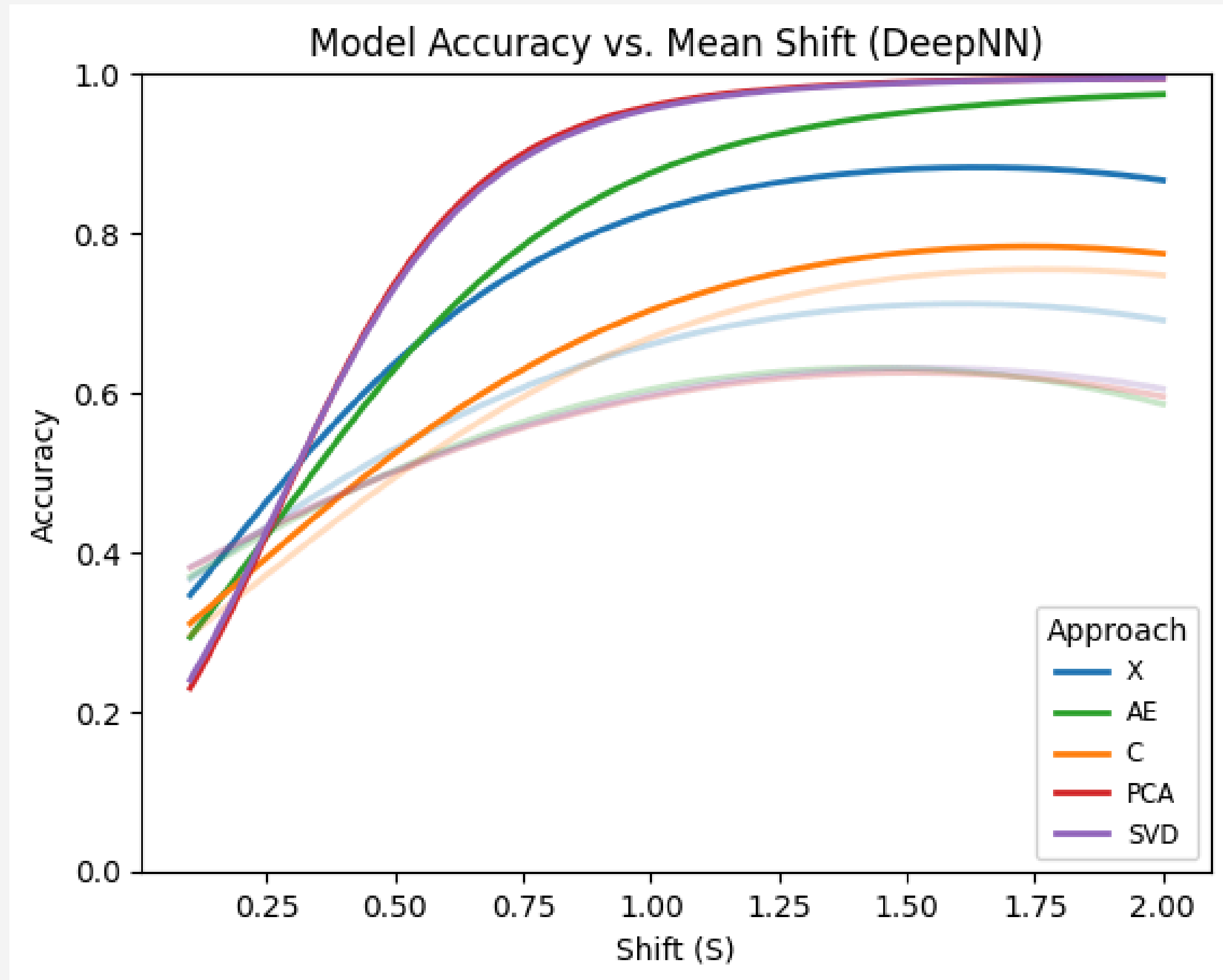


Results - Simulation



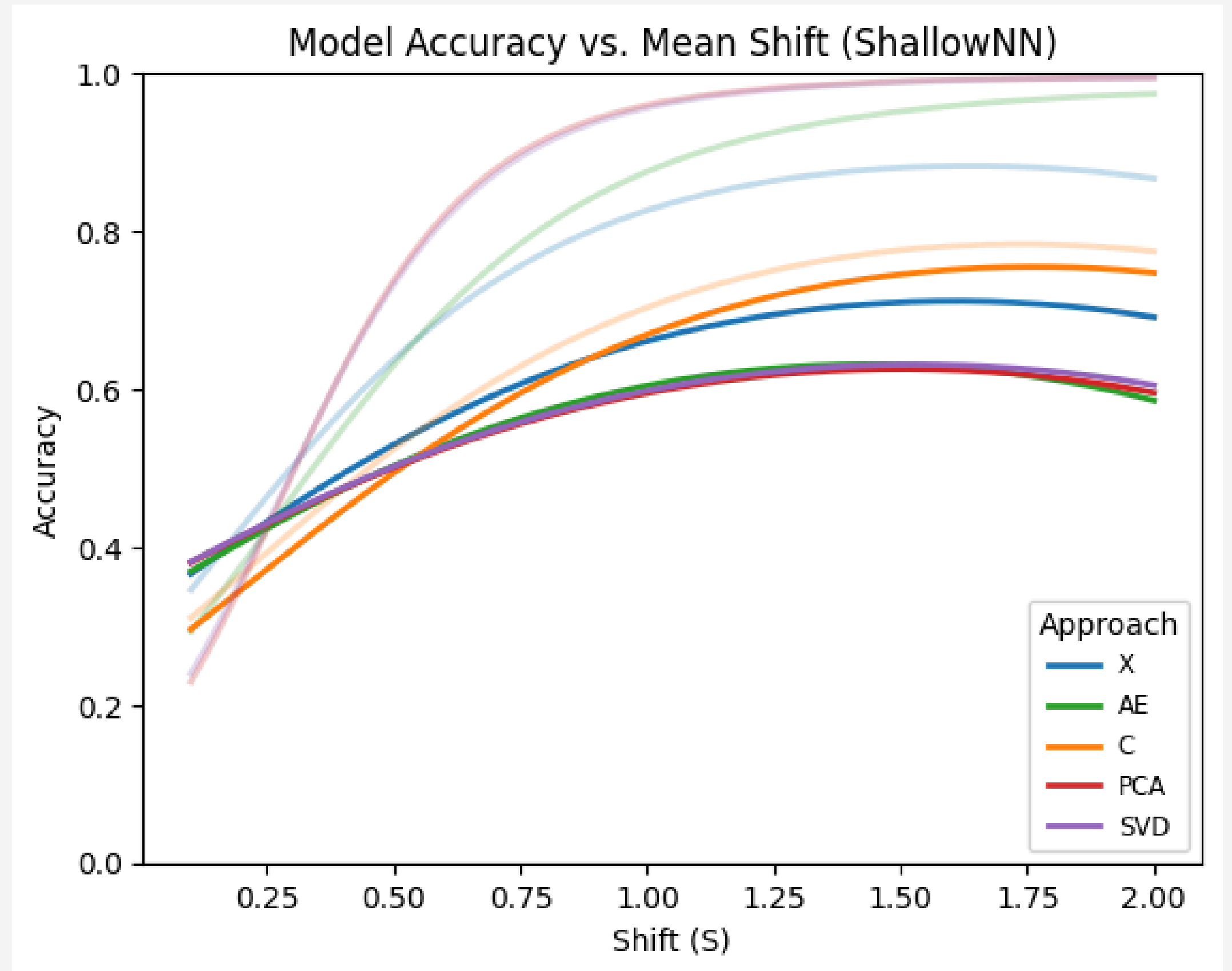
Results

- Simulation



Results

- Simulation



Conclusions

- NMF and Conv didn't perform
- PCA and SVD seems to perform the best
- DeepNN's >> ShallowNN's
- DRT's improve classification accuracy
- Tuned NN's perform slightly better than XGBoost
- Shift correlates with Accuracy

Thank you !

FiN

Simulation

- Based on case data
- Normal Distribution $N(\mu, \sigma)$
- S - mean shift
- P_s - columns with shifted mean
- P_0 - columns with mean 0



$\mu = S * 1$

$\mu = S * 2$

$\mu = S * -1$

$\mu = S * -2$

