WROCŁAW UNIVERSITY
OF ENVIRONMENTAL
AND LIFE SCIENCES

Deep Learning in the bioinformatic modelling of taxonomically annotated microbial communities in aquaculture

Marek Sztuka & Joanna Szyda

Goals:

DL approaches to classification of microbiome data

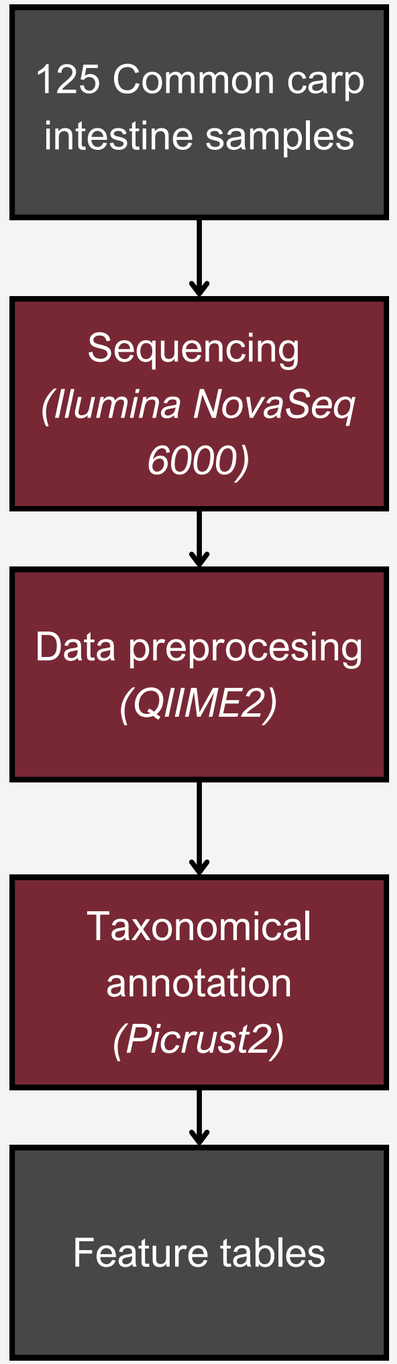
- Architectures (CNN, FNN)
- Dimensionality reduction

Developement of DL model capable of classifing ponds which differed in probiotic supplementation

Materials

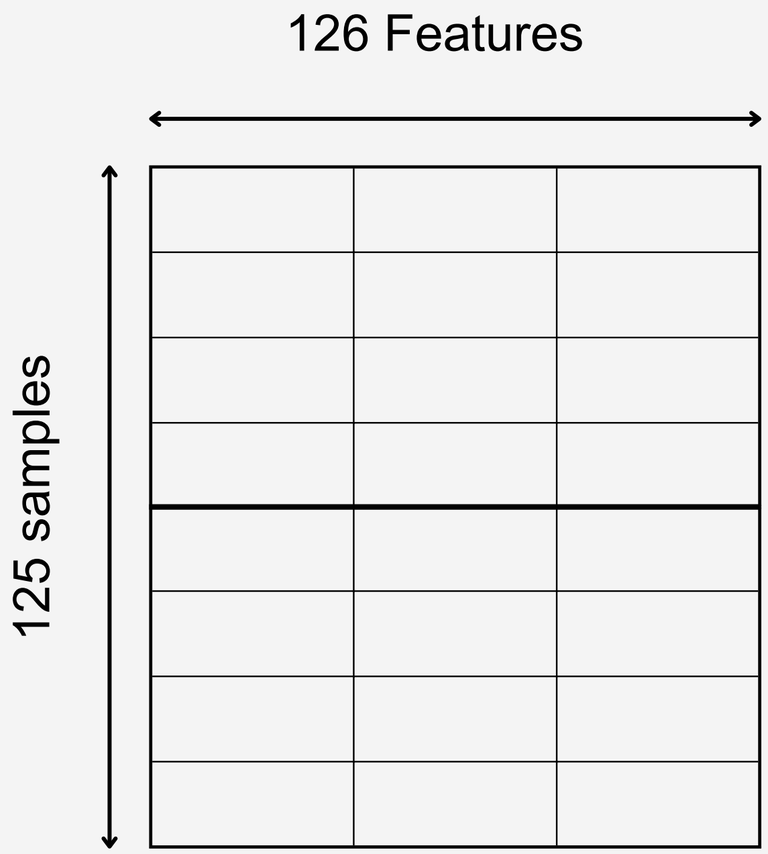
- Enviromental microbiome of fish intestine
- Sequenced reads of 16S rRNA gene
- 5 experimental setups, with different probiotic supplementation

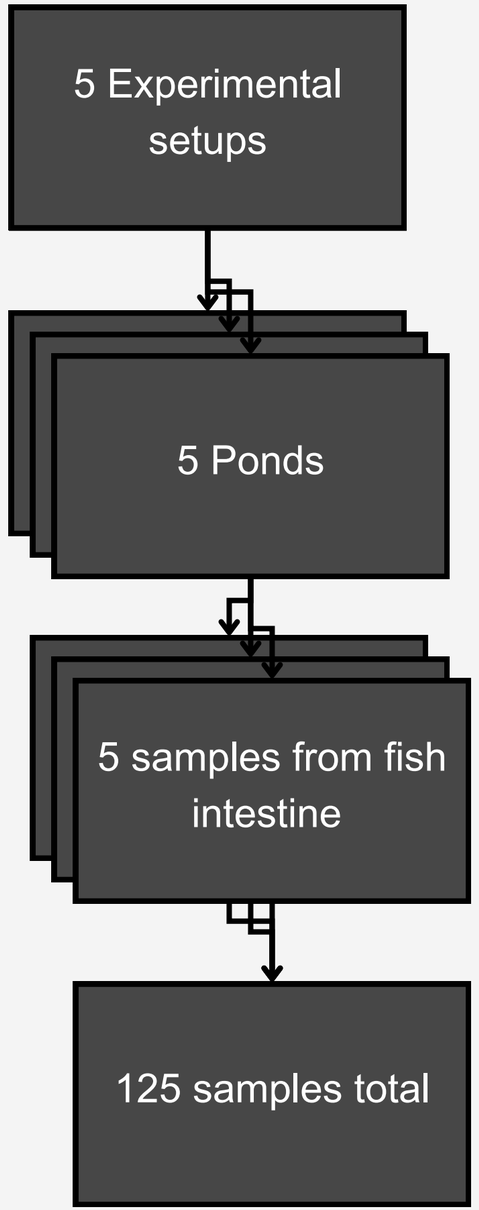
Setup:	Control	1	2	3	4
Supplementation Feed	X	X	EM	X	EM
Enviromental	X	EM farma	EM farma	EM	EM



Materials

- Features - Bacteria families
- Samples - Individuals
- Cells - Bacteria abundance
- Zero inflated data





Input transformation

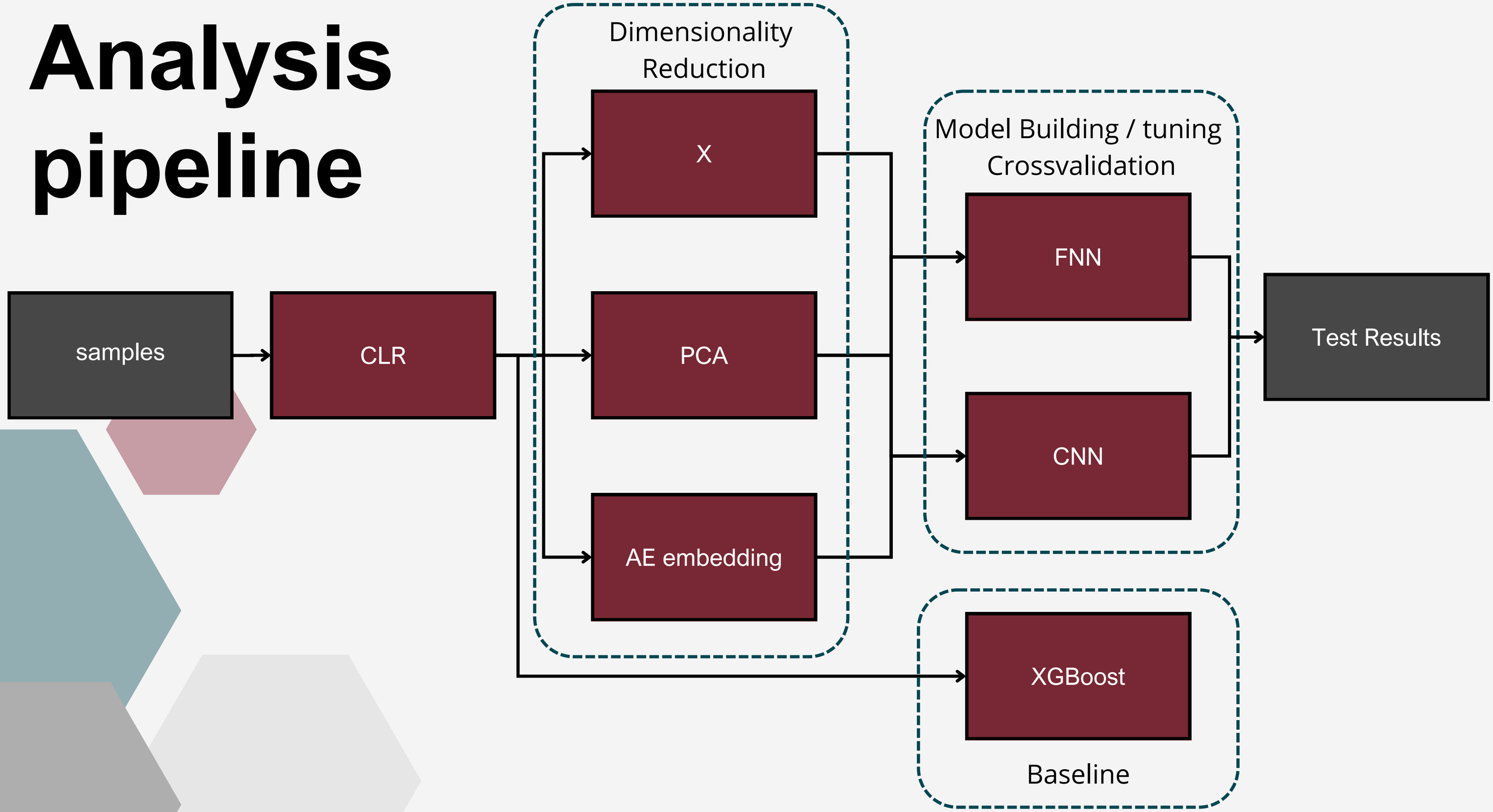
Centered log-ratio transformation

$$CLR(x_i) = ln \frac{x_i}{g(X)}$$

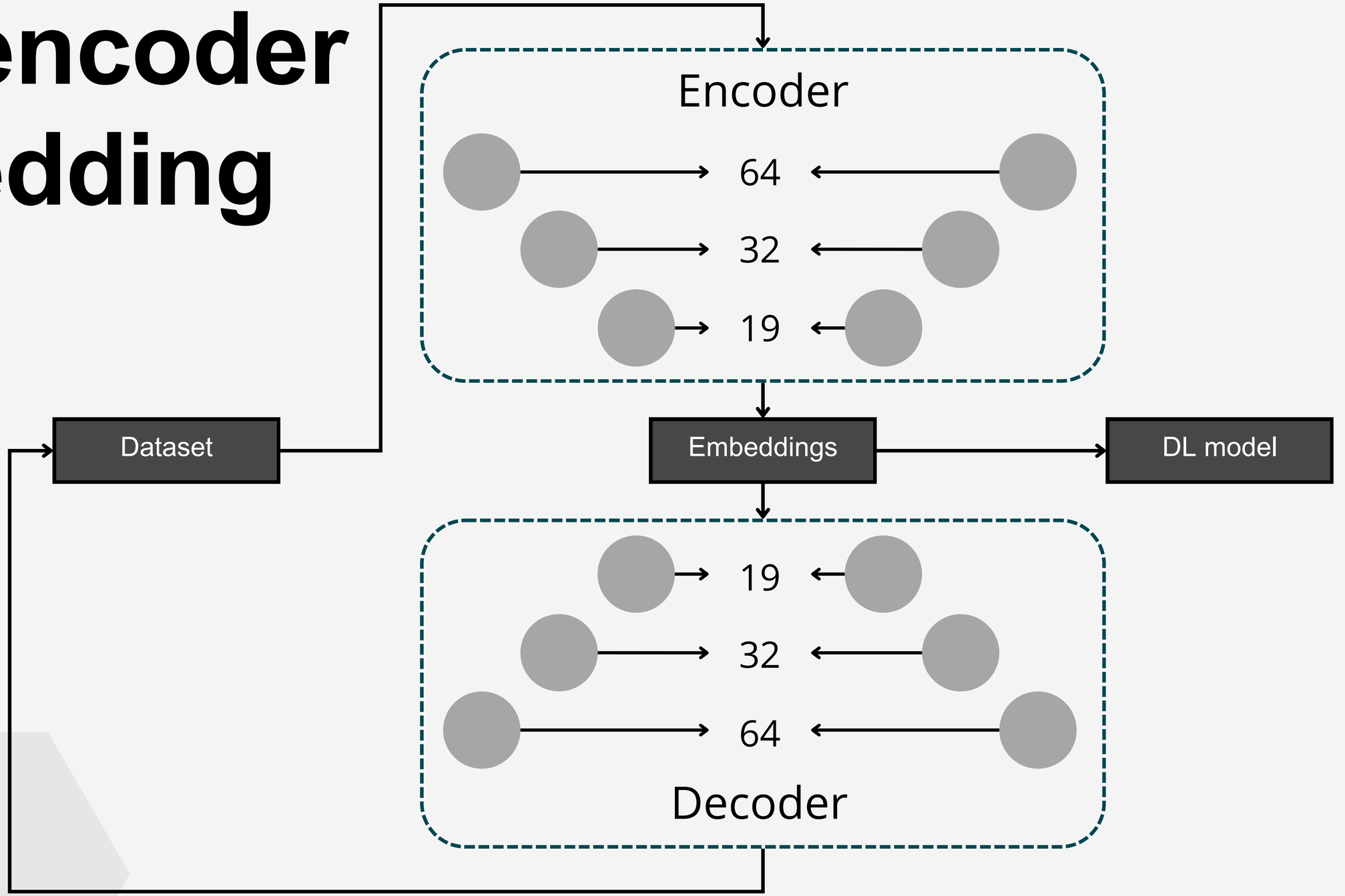
Baseline

- Random class assignment
- XGBoost
  - Depth of the tree: 3
  - eta: 0.1

Analysis pipeline

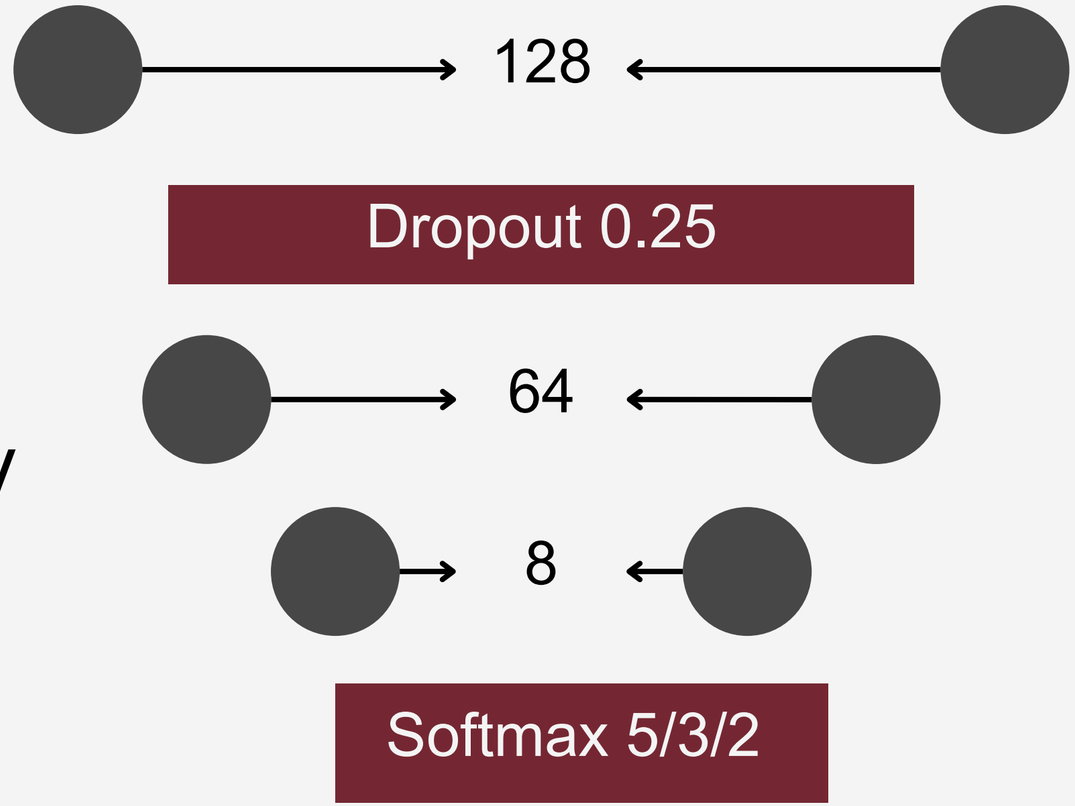


Autoencoder Embedding



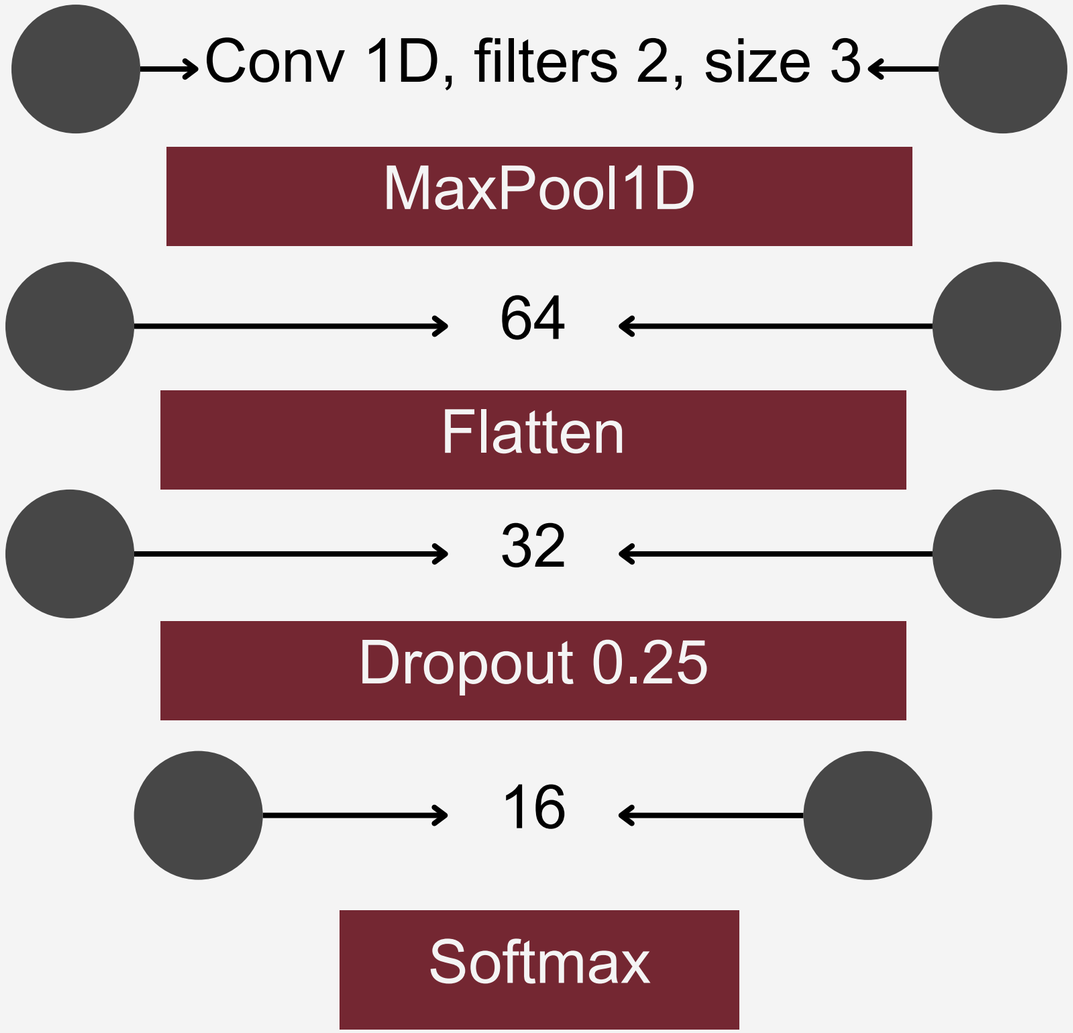
Model Architecture - FNN

- Activation - Relu
- Optimizer - ADAM
- Loss - categorical crossentropy
- Metric - accuracy
- Weight regularizer - L2 (0.01)

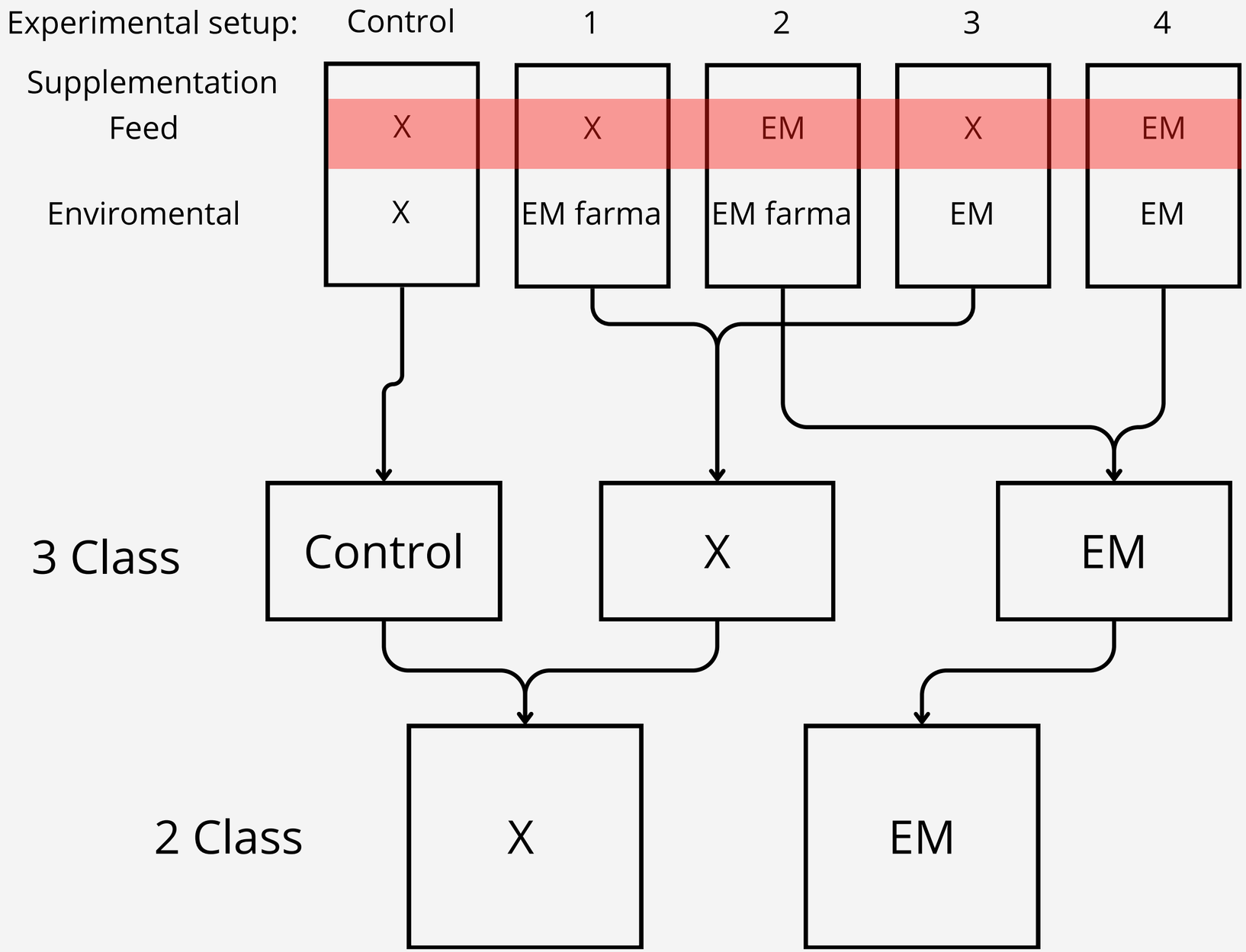


Model Architecture - CNN

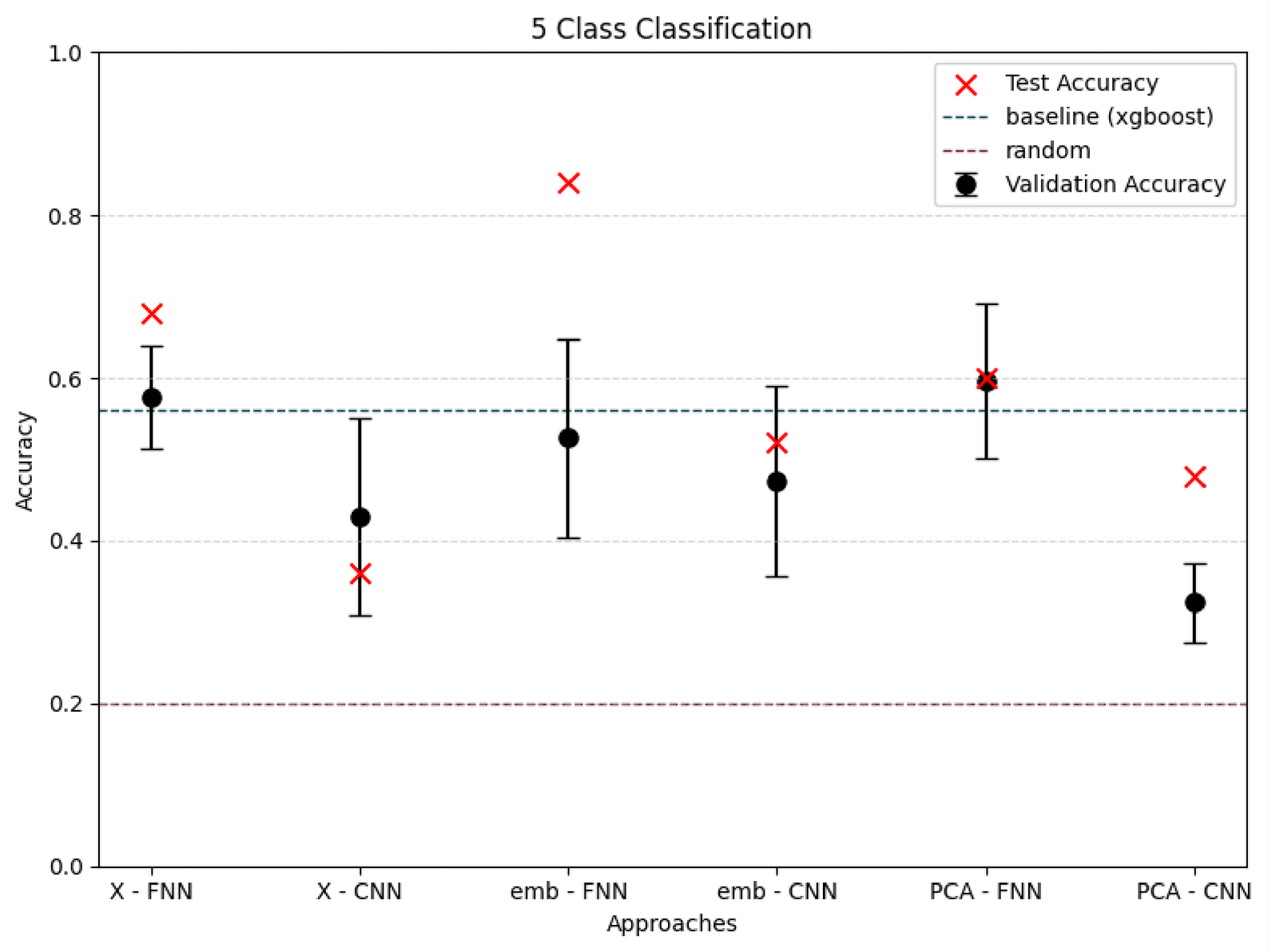
- Activation - Relu
- Optimizer - ADAM
- Loss - categorical crossentropy
- Metric - accuracy
- Weight regularizer - L2 (0.01)



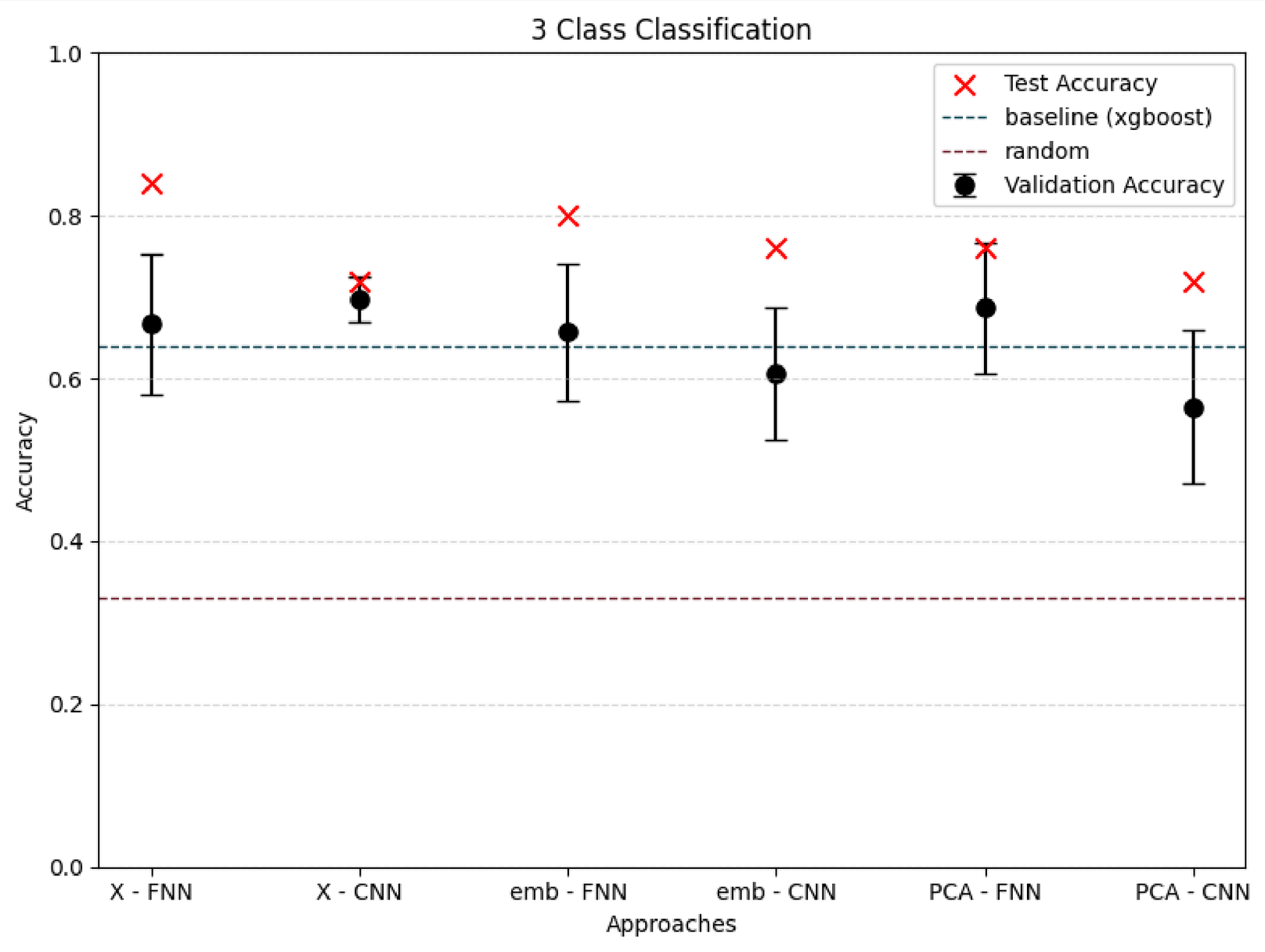
Class reduction



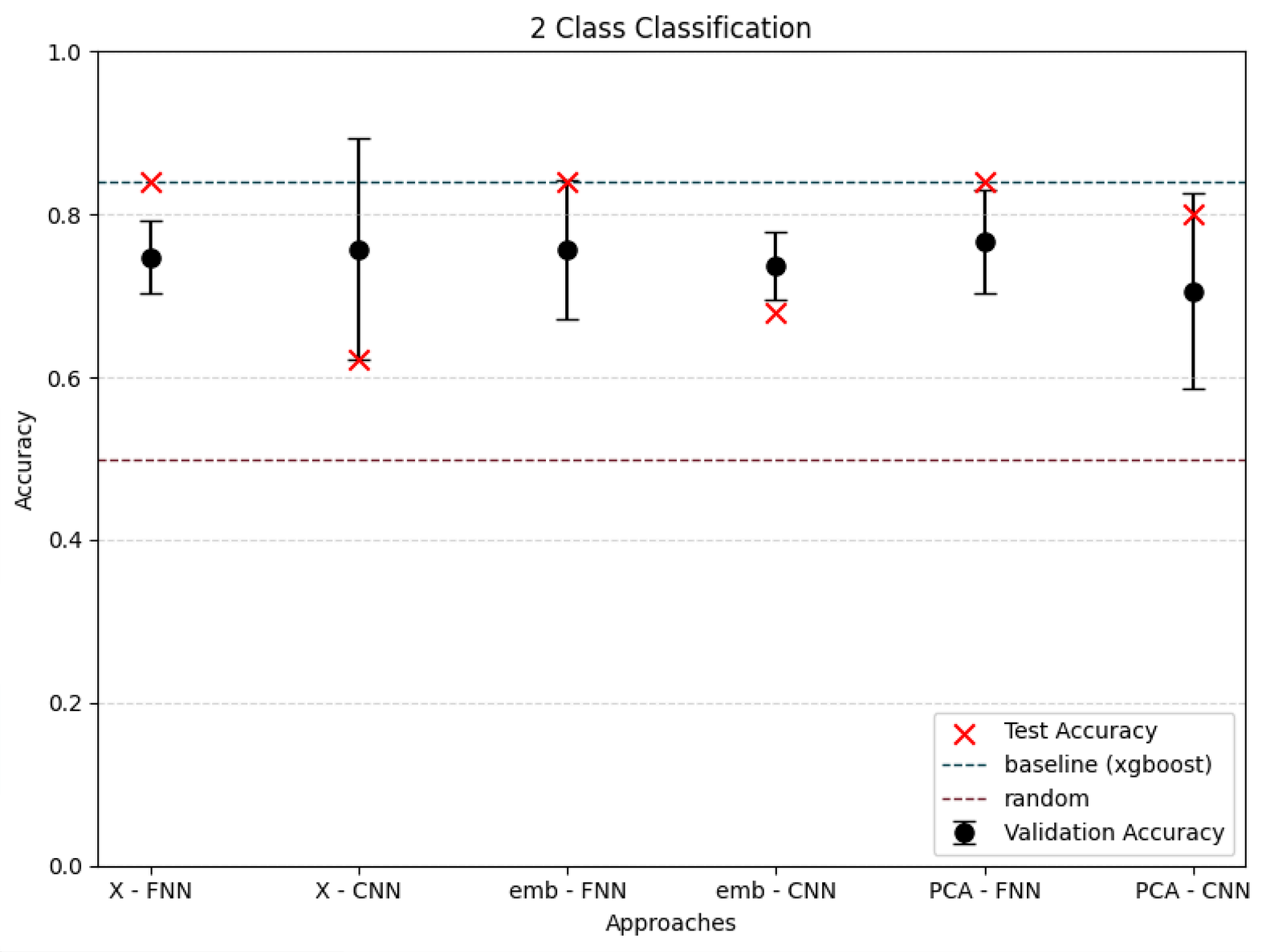
Results 5 class



Results 3 class



Results 2 class



Conclusions Architectures

- Overall increase in accuracy when reducing No. classes
- Worse performance for CNN approaches due to non existent structural patterns
- FNN models proved to be slightly better than baseline xgboost
- Plain Convolution was the worst approach

Conclusions Dimensionality reduction

- Out of dimensionality reduction techiniques embedding yield most satsfactionary results

Fin

Leading Research Group **THETA**

THE BIOSTATISTIC GROUP

LEADER

PROFESSOR JOANNA SZYDA

WROCŁAW UNIVERSITY
OF ENVIRONMENTAL
AND LIFE SCIENCES

BIOSTATISTICS GROUP
WROCLAW, POLAND

Autoencoder Embedding