The Reg challenge

Deep Learning

Jelena Duric
Esteban Aspe
Nicoleta Roman

Process

01

Standarized the dataset

02

Tried HE and Glorot Initializations

03

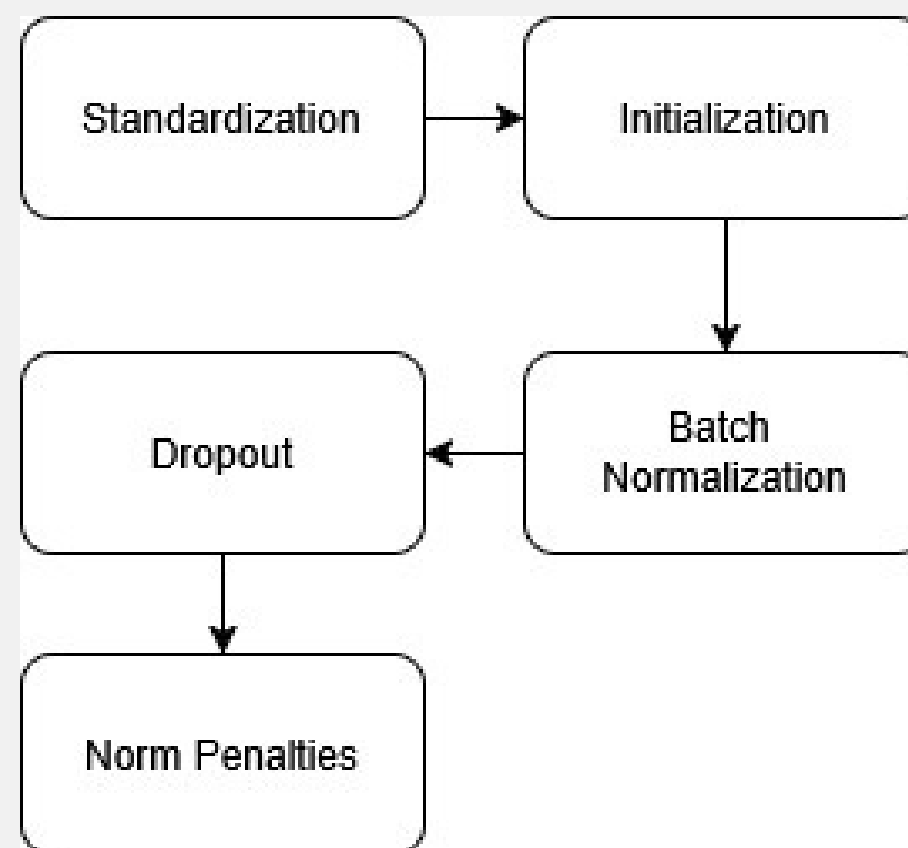Tried with Batch Normalization

04

Tried with different Dropout rates

05

Tried Norm Penalties with different lamdas

06

Tried the best combination in different Architectures

Results after standarization

Given the training surface has been normalized, best Architecture might change

Tried with the best performing Architectures in previous training:

2. 2048-1024
5. **1024-512-256-128**
8. 2048-1024-512-256-128

Architectures after standarization

Legend: ■ Codabenc Accuracy ■ Validation Accuracy
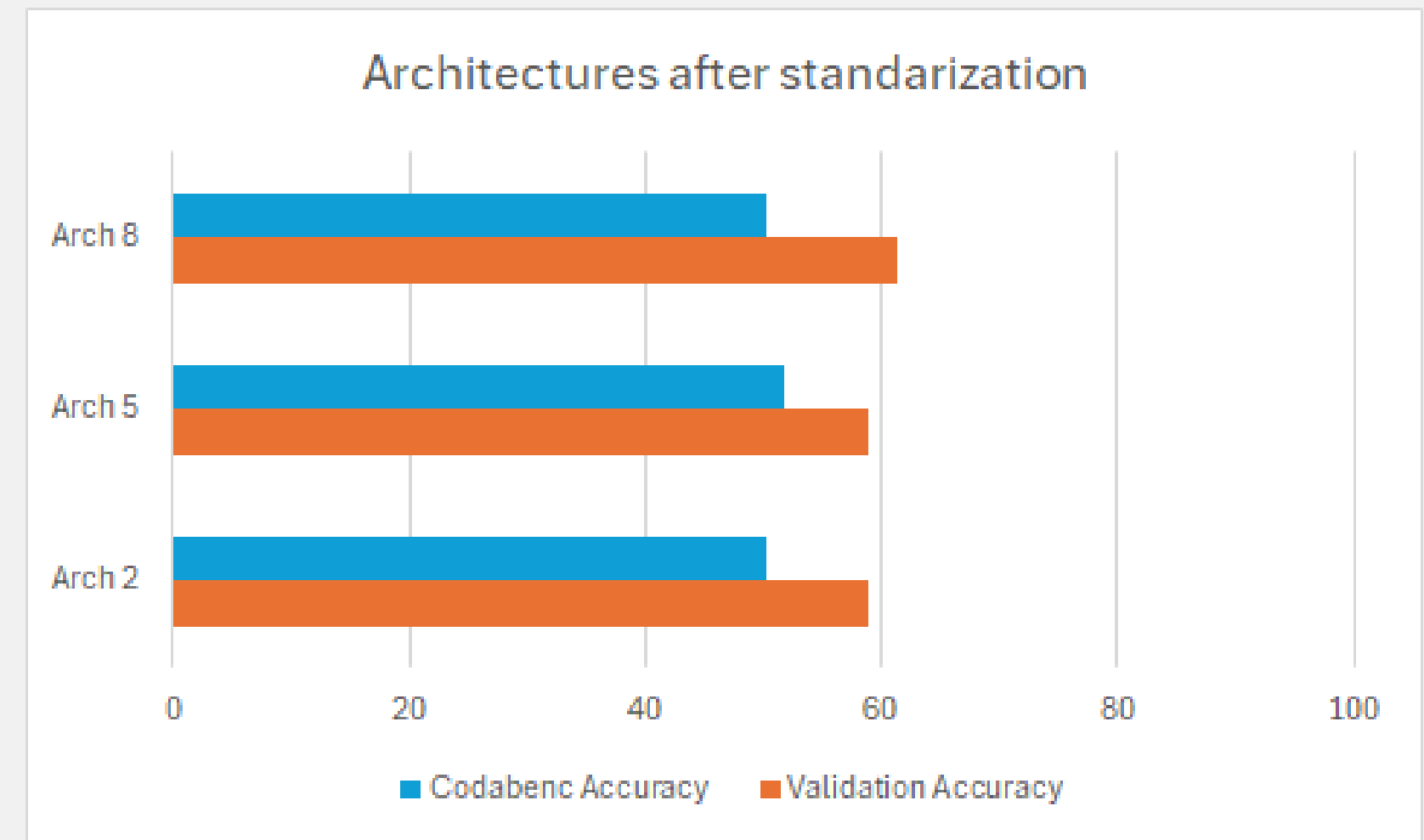
Intialization

Different combinations where tried with Arch 5:
- **HE intitialization for every layer and last one Glorot** with ReLU and **LeakyReLU**
- **Glorot** in **every layer** with **LeakyReLU**. It can not be applied with ReLu since it is designed to keep the variance of the activations constant across layers during training

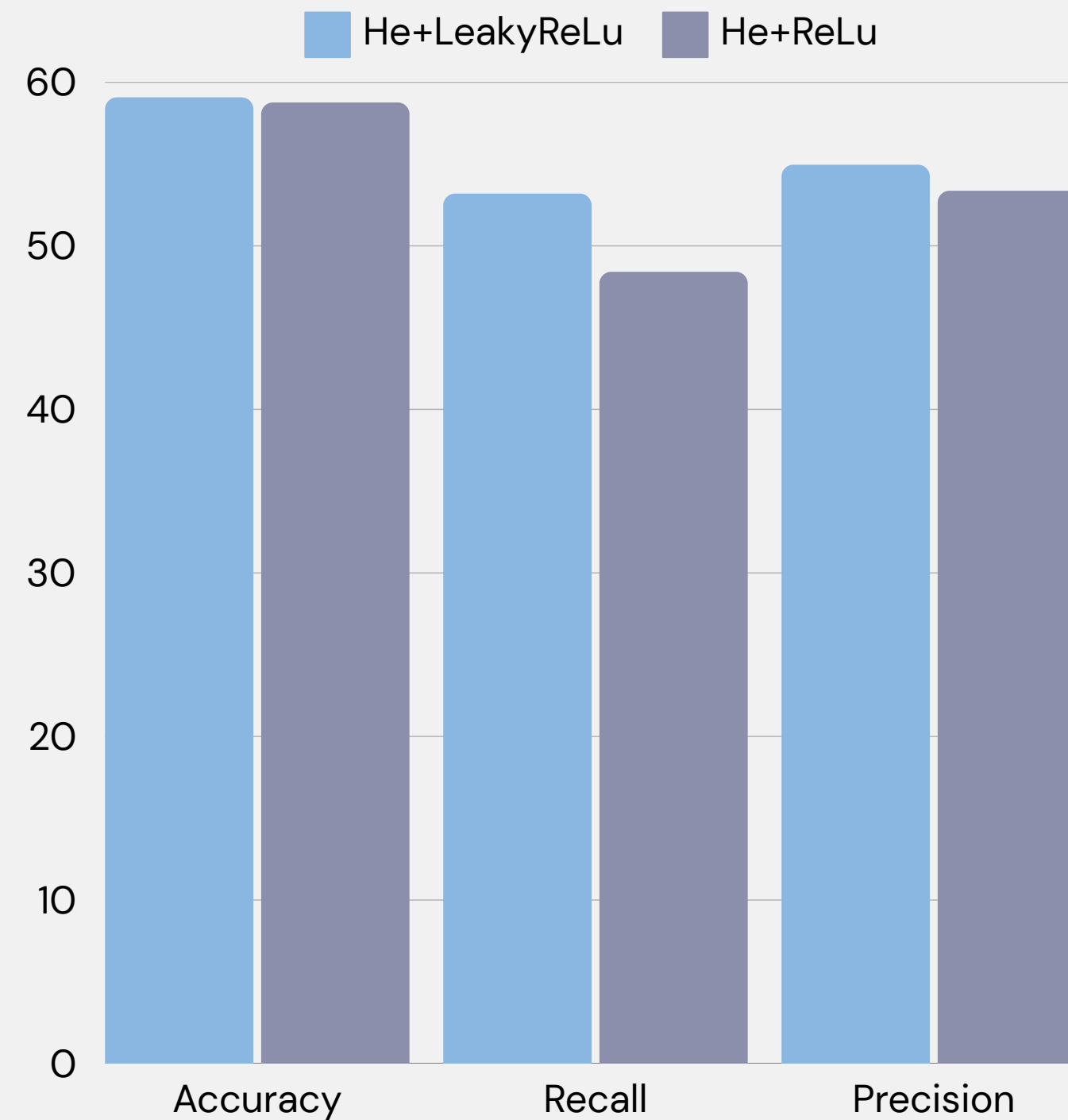# Intialization. He vs Glorot + LeakyReLu
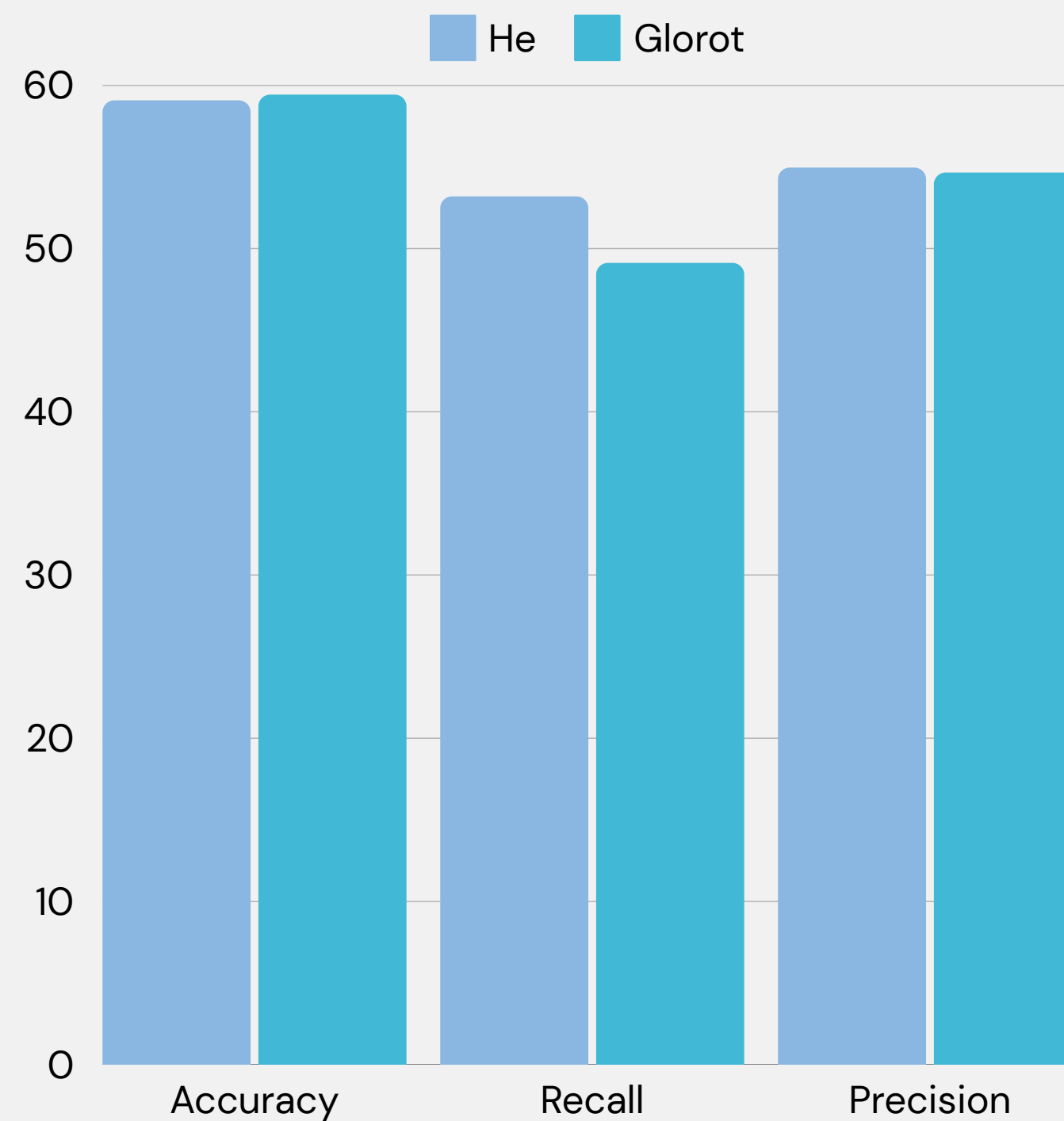
Results:
Network 5 (20 ep) + He + LeakyReLu:
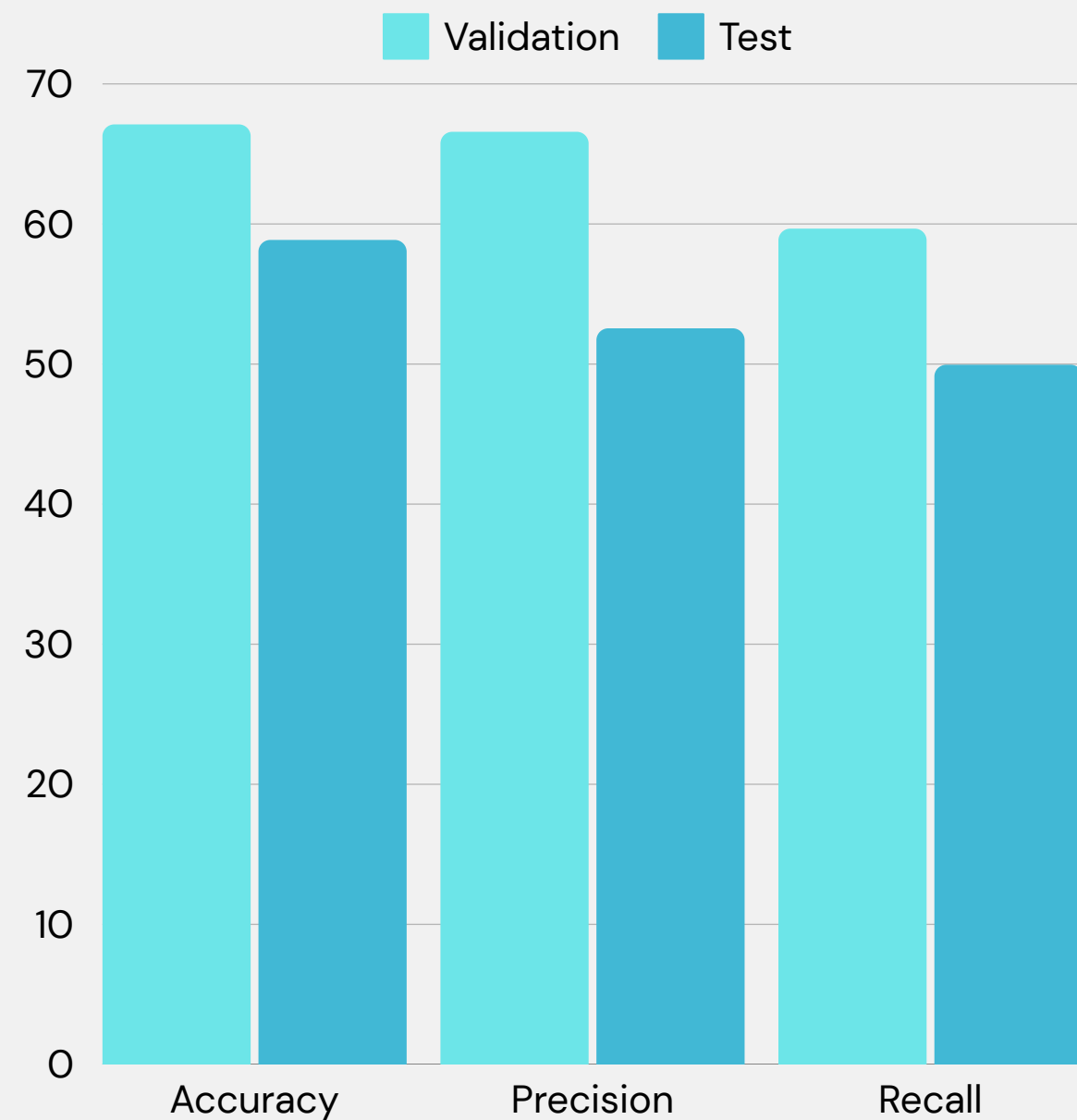
Network 5 (20 ep) + Glorot + LeakyReLu:

Intialization. He + ReLu vs He + LeakyReLu

Results:
Network 5 (20 ep) +
He +ReLu

Network 5 (20 ep) +
He + LeakyReLu

Legend: He+LeakyReLu, He+ReLu

Chart categories: Accuracy, Recall, Precision

Y-axis: 0, 10, 20, 30, 40, 50, 60

Batch Normalization

*Best result (batch norm. before)

- Batch normalization was tried **before** and **after** activation function (for network 5, 100 epochs, He initializer, LeakyReLu, Dropout 50->20).
- No significant difference was noticed, however applying it **before** was slightly **better.**
- Additionally, the applying the natch norm. before was also tried with **architecture 2 (1024-512-256)** 100 epochs, He initializer, LeakyReLu, Dropout 50->30 and similar results were obtained with **no significant differences**.

Dropout and Norm Penalties

- Two dropouts were tried with Arch 5, **50–40–30–20** and 40–30–35–45. Both prevented overfitting with 20 epochs. With 60 epochs it overfitted, but the validation accuracy improved.
- A third dropout **20–30–40–50** was tried with Arch 5, but **presented no difference** compared to 50–40–30–20 (both tried for 100 epochs).

- Ridge Regularization technique was used, and it partially prevented overfitting. However, dropout got better results.

- The combination of both techniques produced underfitting in most of the combinations tried. Further research would be needed.
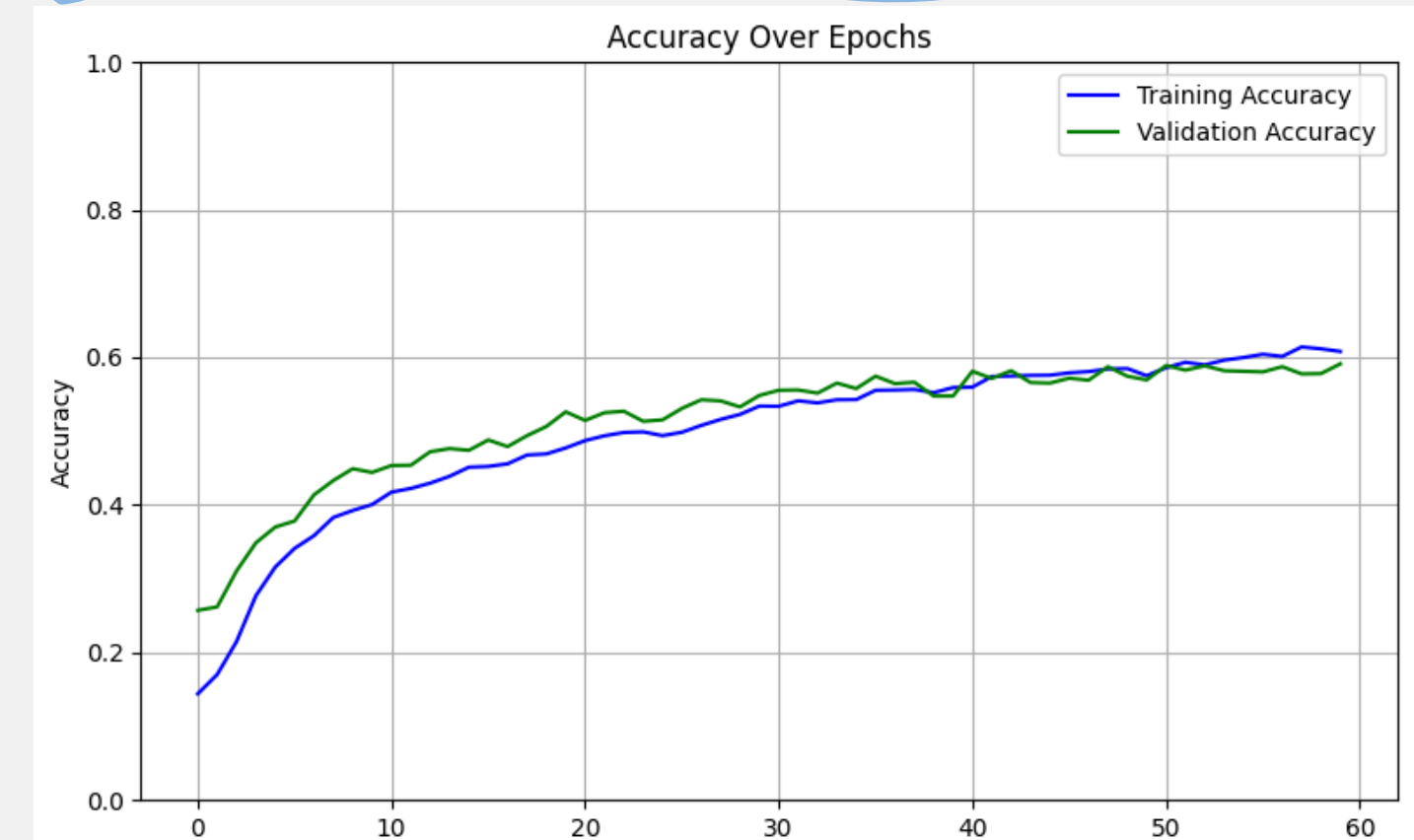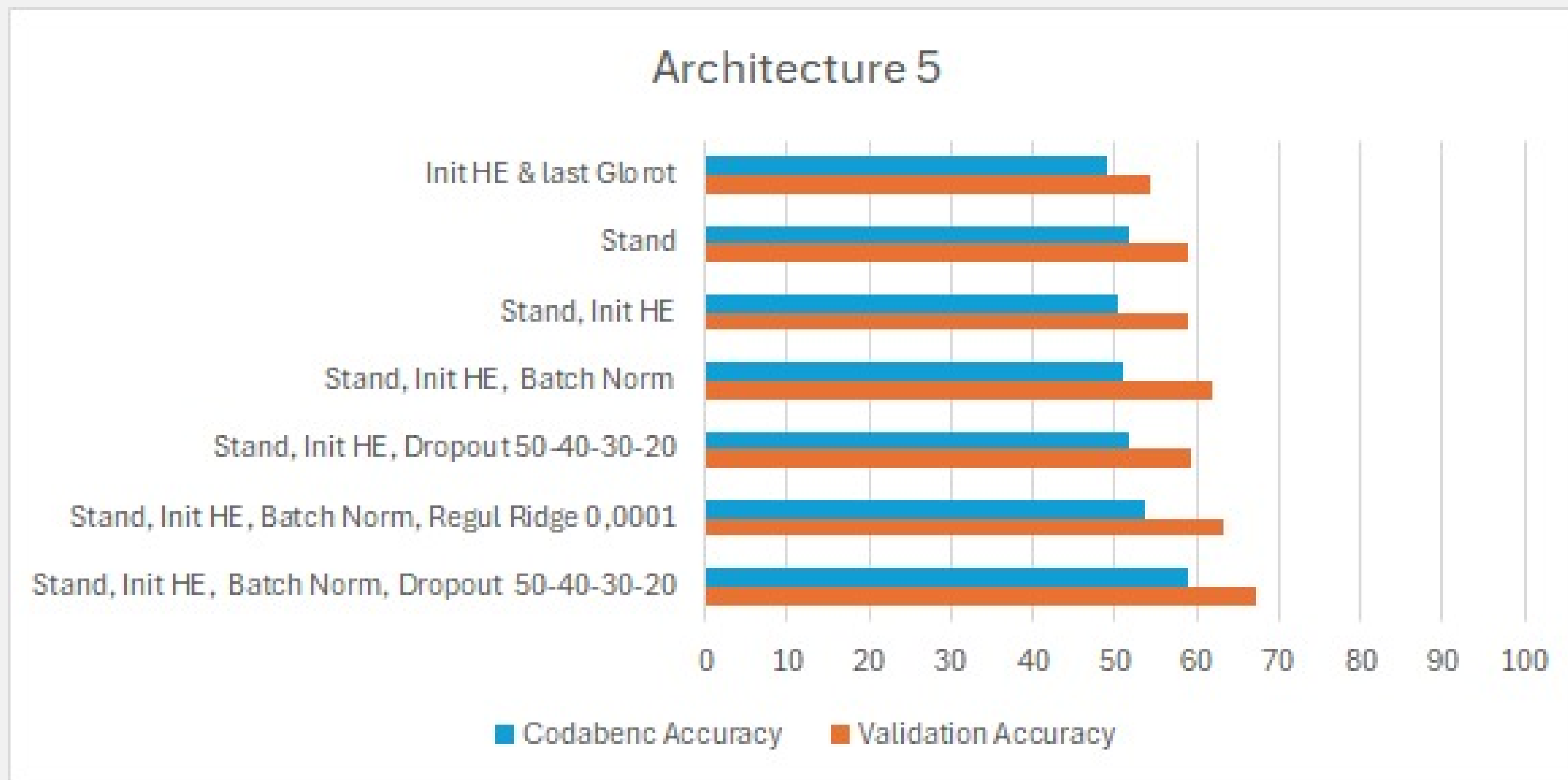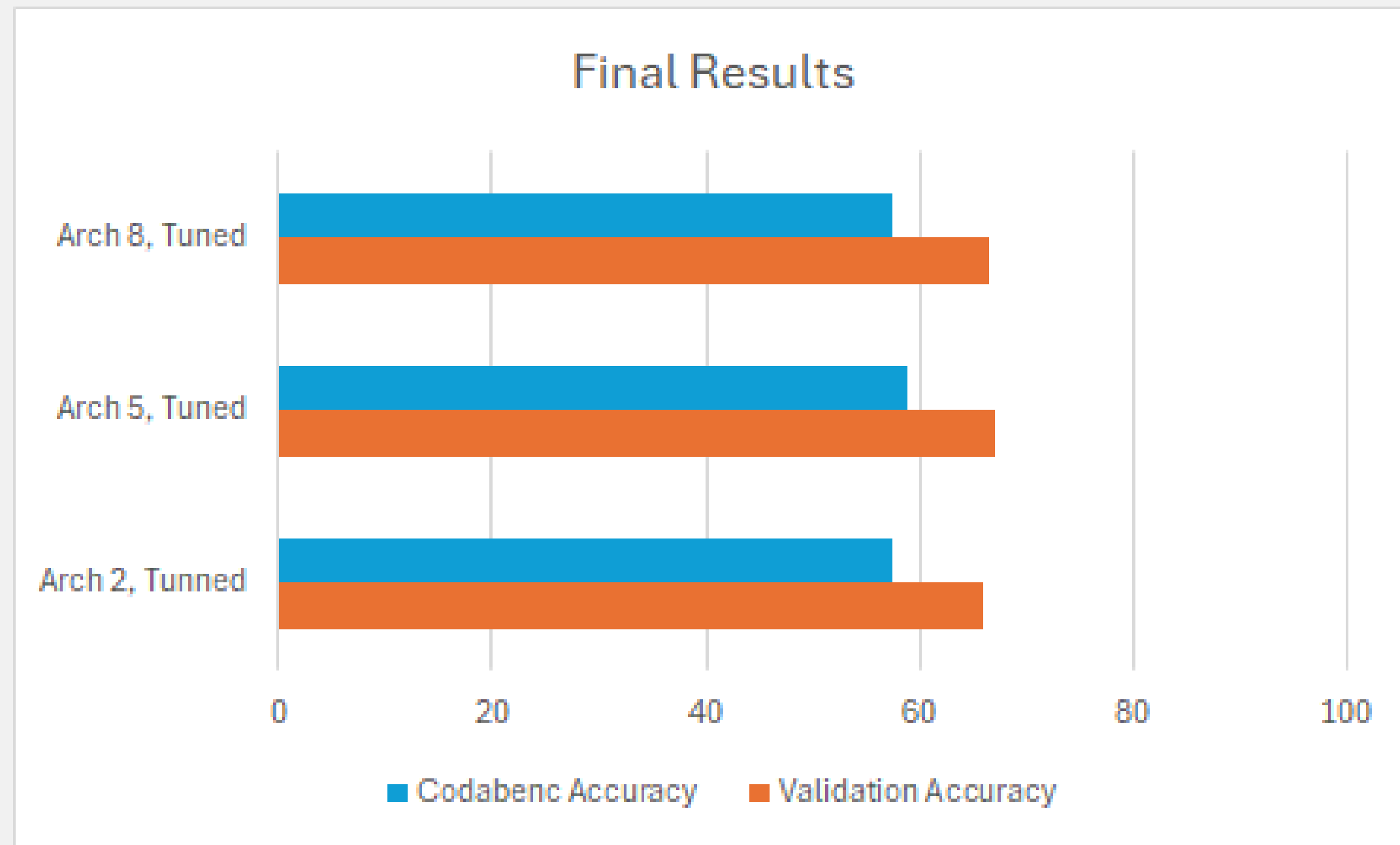

Dropout
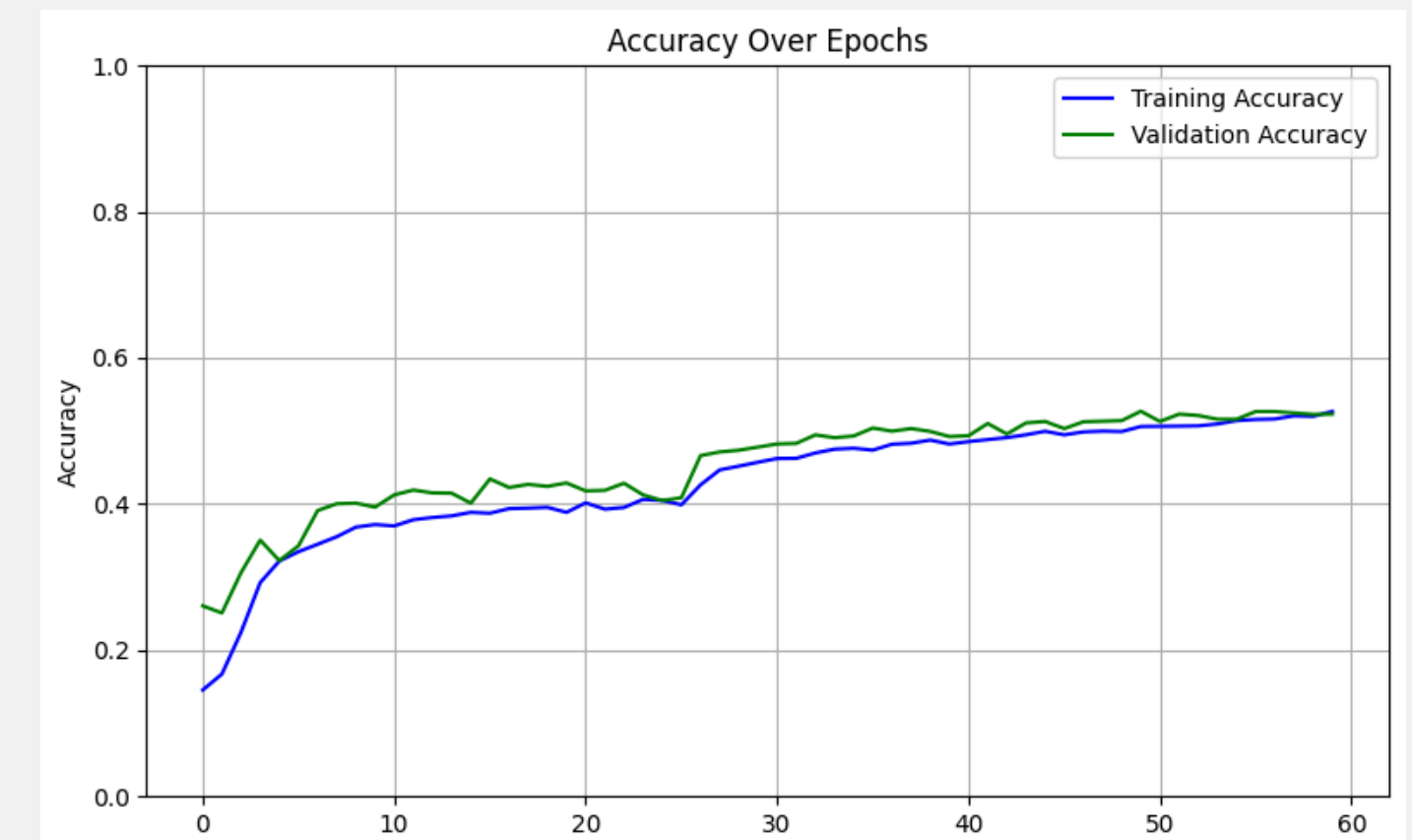
Ridge

Discussion. Comparison
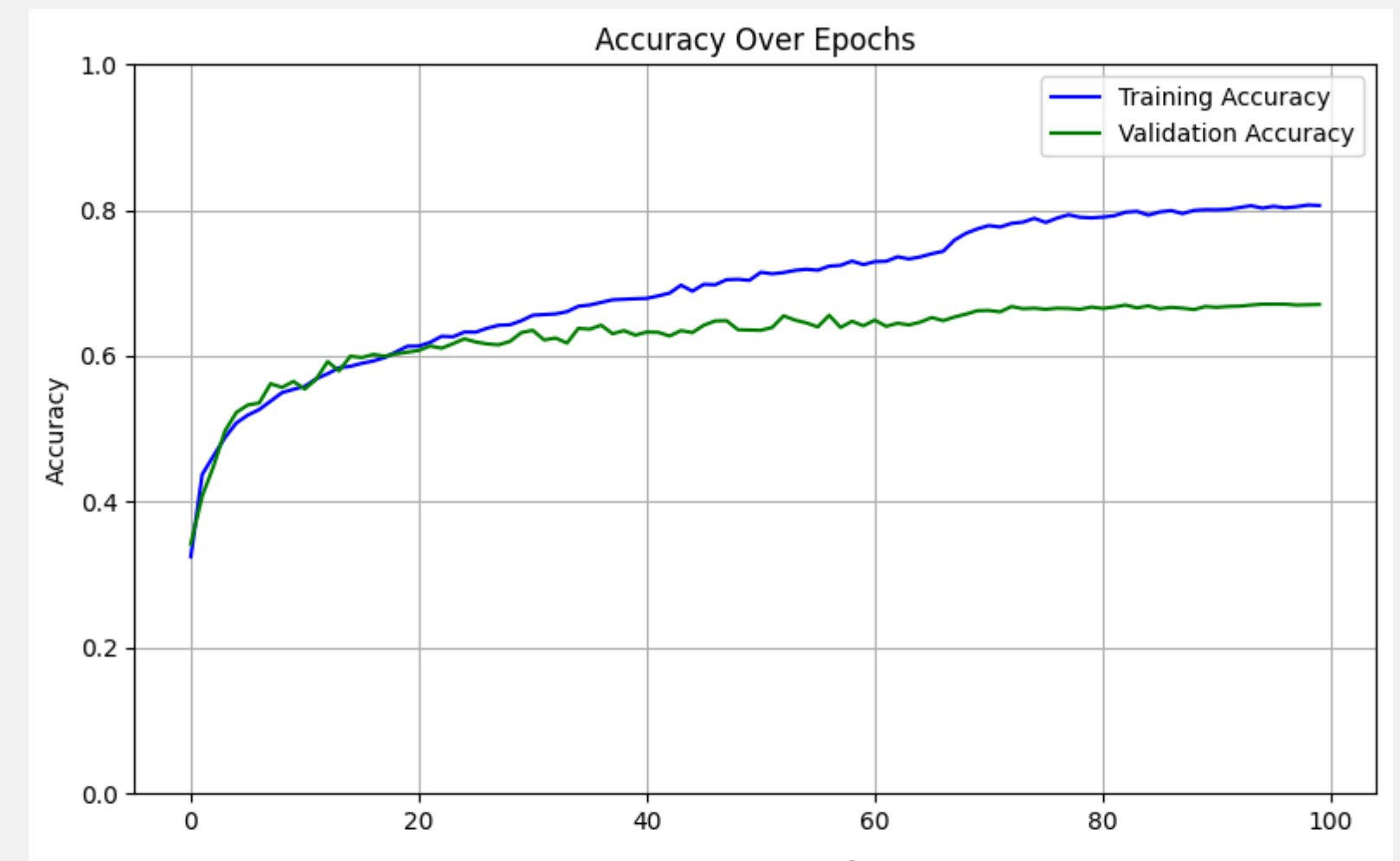
Discussion. Comparison

Conclusions

The **best performing neural networks** were network 5 and 8 with 100 epochs, standarized dataset, HE initialization and Glorot for last layer, Batch Normalization before the act. function and Dropout 50-40-30-20, 60-50-40-30-20, respectively

Network	Mean Accuracy	Mean Precision	Mean Recall
5	58.86	52.55	49.95
8	57.5	59.16	57.17

Network 5

Thank you