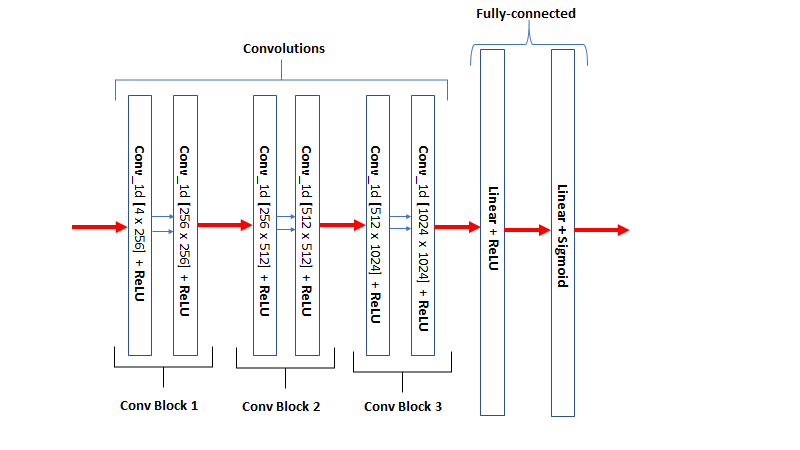


The following is a summary of the hyper parameter optimisation tests after 10 epochs of training and testing:

**Selected CNN Architecture for these experiments:**



**Summary of the different configurations tested for all 4 categories:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Optimiser*** | ***Batch Size*** | ***Dropout*** | ***L1*** |
| ***exp1*** | *SGD* | *256, 256* | *0.3, 0.3* | *e-12* |
| ***exp2*** | *SGD* | *256, 256* | *0.3, 0.3* | *e-5* |
| ***exp3*** | *Adam* | *256, 256* | *0.3, 0.3* | *e-12* |
| ***exp4*** | *SGD* | *256, 256* | *0.2, 0.4* | *e-12* |
| ***exp5*** | *SGD* | *128, 128* | *0.2, 0.2* | *e-12* |
| ***exp6*** | *SGD* | *256, 256* | *0.3, 0.3* | *e-16* |
| ***exp7*** | *SGD* | *256, 256* | *0.3, 0.5* | *e-12* |
| ***exp8*** | *SGD* | *64, 64* | *0.3, 0.5* | *e-12* |
| ***exp9*** | *SGD* | *16, 16* | *0.3, 0.5* | *e-12* |

**Optimisation SUMMARY:**

The tests show that changing the selected hyper parameters for this optimisation problem did vary the model’s performance. Having stated that, the model responded to a greater extent to certain parameters than others for different data categories. For example, lowering the batch sizes had different effects on the data, with EpiMap responding negatively to lower batch sizes, as opposed to HumanFC where the model did improve with lower batch sizes (to a certain degree). The optimiser tests could have been done more thoroughly, as TF responded positively to ADAM optimiser and outperformed all other SGD tests.

Certain parameters have been identified to be more influential than other for the different categories and show a more precise framework to tweaking the hyper parameters to allow for more meaningful optimisation experiments, and ultimately, find the best configurations that perform the best for each data category.

Another note, the decision to split the model from a super model for all 4 data categories, to 4 individual models for each category has paid off. Building 4 separate models that can be tweaked and optimised individually as opposed to training one super model is easier to interpret but does come at a cost of computational efficiency.

-

- **Cage:**

Experiment 7 configuration appears to be the most appropriate model (so far) for Cage, even though it is not superior to the others. Its TPR score is the highest (0.74) but does not show the lowest Loss core. Surprisingly, the worst performing model in terms of TPR scores (experiment 3) had the lowest loss score even after 10 epochs of training. This suggests that the best architecture is yet to be found for this category and further tests should be carried out.

Out of these configurations, ***Experiment 7*** would be the best candidate for Further chromosome and model comparison testing.

*Experiment 7 :*

|  |  |  |  |
| --- | --- | --- | --- |
| **Optimiser** | **Batch Sizes [Train, Test]** | **Dropout [After Conv block 1, After Conv Block 2]** | **L1** |
| SGD | [256, 256] | [0.3, 0.5] | e-12 |

**- HumanFc:**

Experiment 8 shows the lowest loss score by a considerable margin, although still relatively high in terms of optimal loss measurements. Unlike Cage, the lowest loss configuration ALSO has the highest accuracy (0.66), indicating that having smaller sized batches for this category results in better TPR scores and lower loss values (up until a ‘’particular” minimum, as exp 9 shows it begins to degrade with even less batch sizes). Whether or not this is global or local minimum, this cannot be told, and further tests will need to be run with lower batch sizes and varying other parameters.

Out of these configurations, ***Experiment 8*** would be the best candidate for Further chromosome and model comparison testing.

*Experiment 8:*

|  |  |  |  |
| --- | --- | --- | --- |
| **Optimiser** | **Batch Sizes [Train, Test]** | **Dropout [After Conv block 1, After Conv Block 2]** | **L1** |
| SGD | [64, 64] | [0.3, 0.5] | e-12 |

**- Epimap:**

Experiment 7 again had the highest TPR score but not the lowest loss score. Comparing this with the lowest loss configuration (experiment 9), the margin is not that great, and the trade-off will be a 20% drop in TPR scores.

Out of these configurations, ***Experiment 7*** would be the best candidate for Further chromosome and model comparison testing.

*Experiment 7 :*

|  |  |  |  |
| --- | --- | --- | --- |
| **Optimiser** | **Batch Sizes [Train, Test]** | **Dropout [After Conv block 1, After Conv Block 2]** | **L1** |
| SGD | [256, 256] | [0.3, 0.5] | e-12 |

**- TF:**

Tf ‘s results were unique, with the only test that involved a different optimiser function (Experiment 3 which used ADAM) performing the best. Lowest Loss score, highest TPR, but still very low TPR scores, thus suggesting that Stochastic Gradient descent may not be the optimal option for this category and further tests using ADAM may be required.

Out of these configurations, ***Experiment 3*** would be the best candidate for further chromosome and model comparison testing.

*Experiment 3:*

|  |  |  |  |
| --- | --- | --- | --- |
| **Optimiser** | **Batch Sizes [Train, Test]** | **Dropout [After Conv block 1, After Conv Block 2]** | **L1** |
| ADAM | [256, 256] | [0.3, 0.3] | e-5 |

**Chromosome-wise Tests and model comparisons:**

* **Epimap**
  + **Optimised DeepPsych model**
  + **DeepSea model**
  + **RNN model**
* **HumanFc**
  + **Optimised DeepPsych model**
  + **DeepSea model**
  + **RNN model**
* **Cage**
  + **Optimised DeepPsych model**

* + **DeepSea Model**
  + **RNN Model**
* **TF**
  + **Optimised DeepPysch Model**

* + **DeepSea Model**
  + **RNN Model**