# Analysis & Prediction of Efficiency using Machine Learning

Include ML in the introduction part.

Write a comparative study on Machine Learning Algorithms used focusing on their results and loss scores.

Write a comparative study on Deep Learning architectures used focusing on their results and loss scores.

**Pre-processing Dataset**

1. Encoding Morphology (0D, 1D, 3D)

2. Encoding Dye using Label Encoder (we get the number of unique items in dye column and assign a numeric value to each of the item present in the dye column)

2. Taking all independent variables/inputs on X-axis by dropping efficiency column and assigning the dependent variable/output/prediction value to Y-axis i.e efficiency.

4. Spliting X (morphology,dye,Jsc,Voc,FF,Passivation) & Y(efficiency) into x_train,y_train for training and x test, y test for testing. We are taking 10% of the dataset for testing

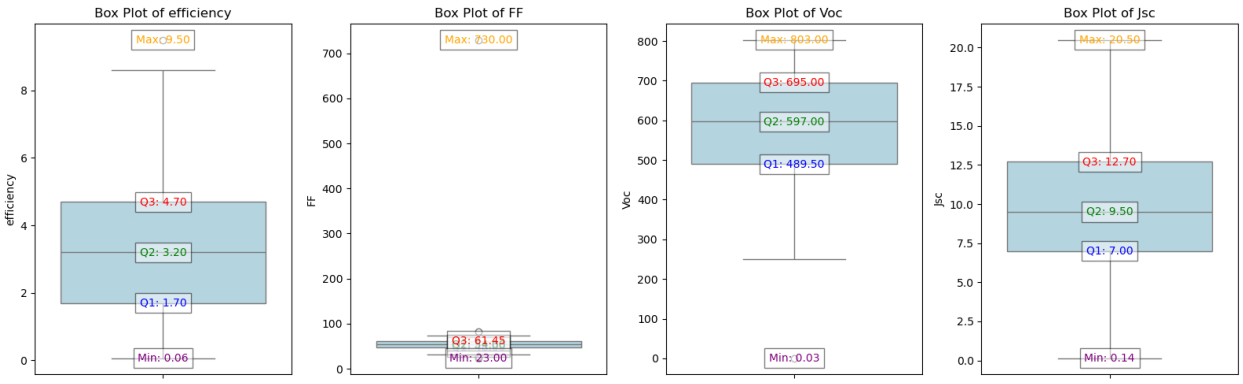
X_train consists of 147 rows and 6 columns(morphology,dye,Jsc,Voc,FF,Passivation)

Y_train consists of 147 rows and a single column(efficiency)

Xtest consists of 17 rows and 6 columns and Ytest consists of 17 rows and 1 column

4*(optional step done) We check for NaN values/ Empty values in the dataset and remove those for final cleaning of the dataset

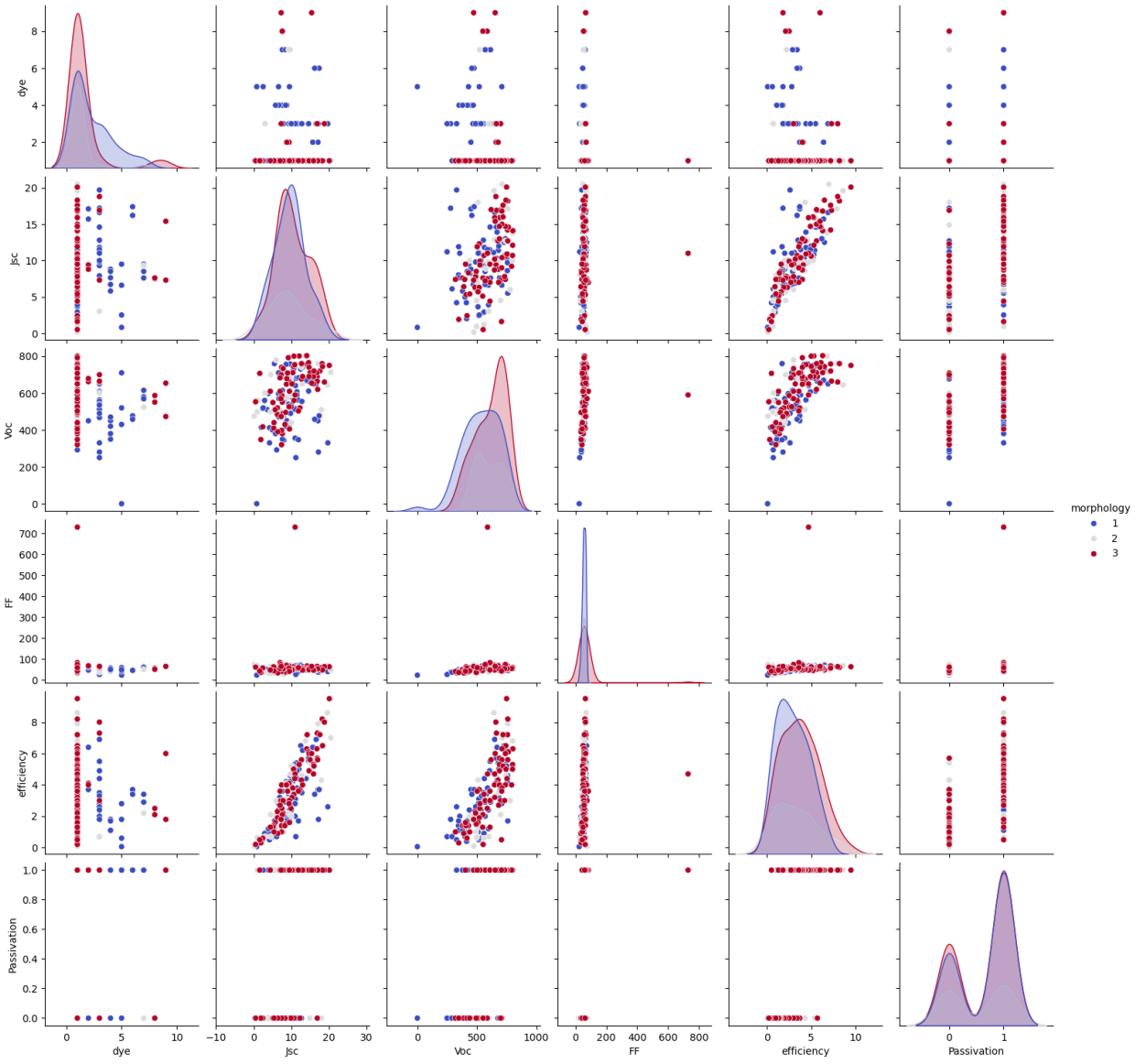
5. From the box plot we see that FF & Voc are very high in value so we need to normalize



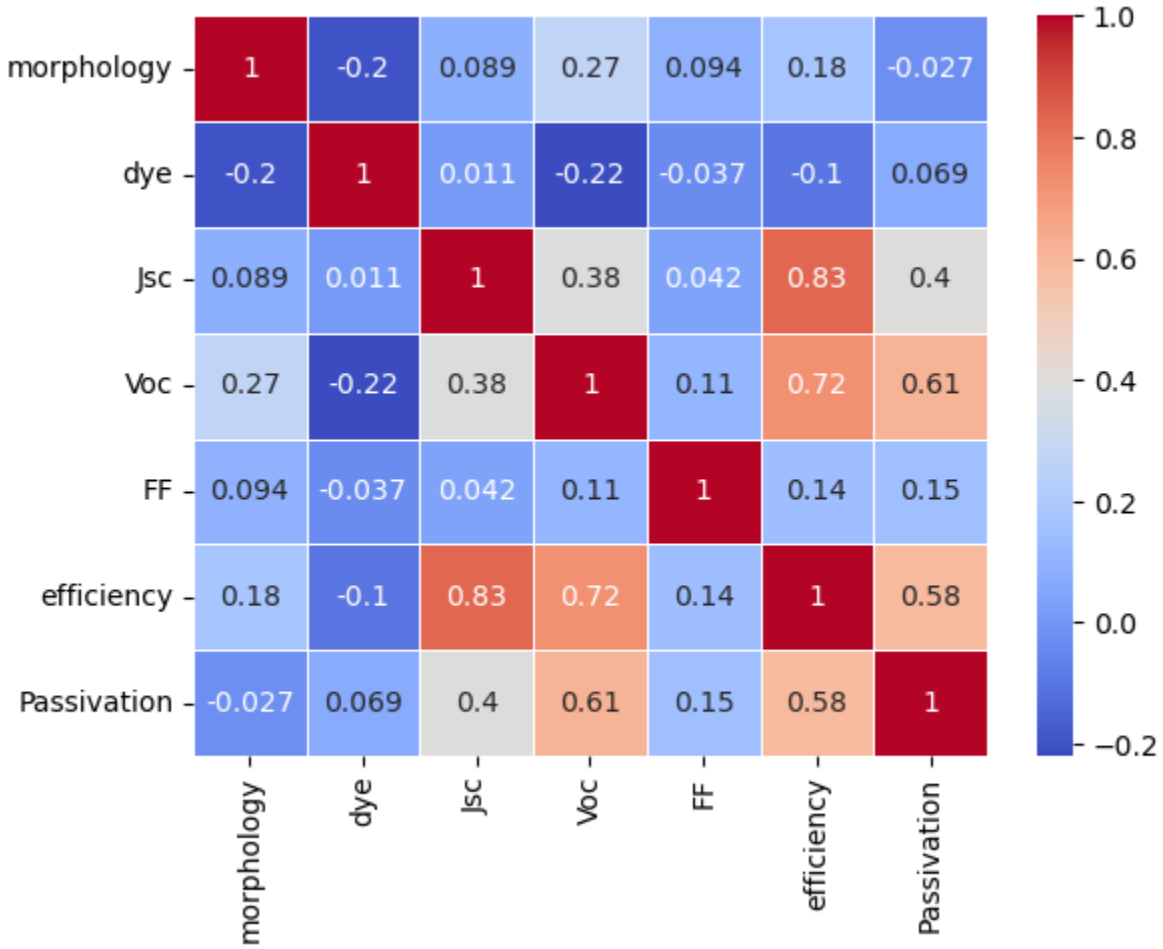
6. Distribution :

**scatterplot matrix**. It shows pairwise relationships between the features in your dataset, with each individual plot displaying the relationship between two variables.

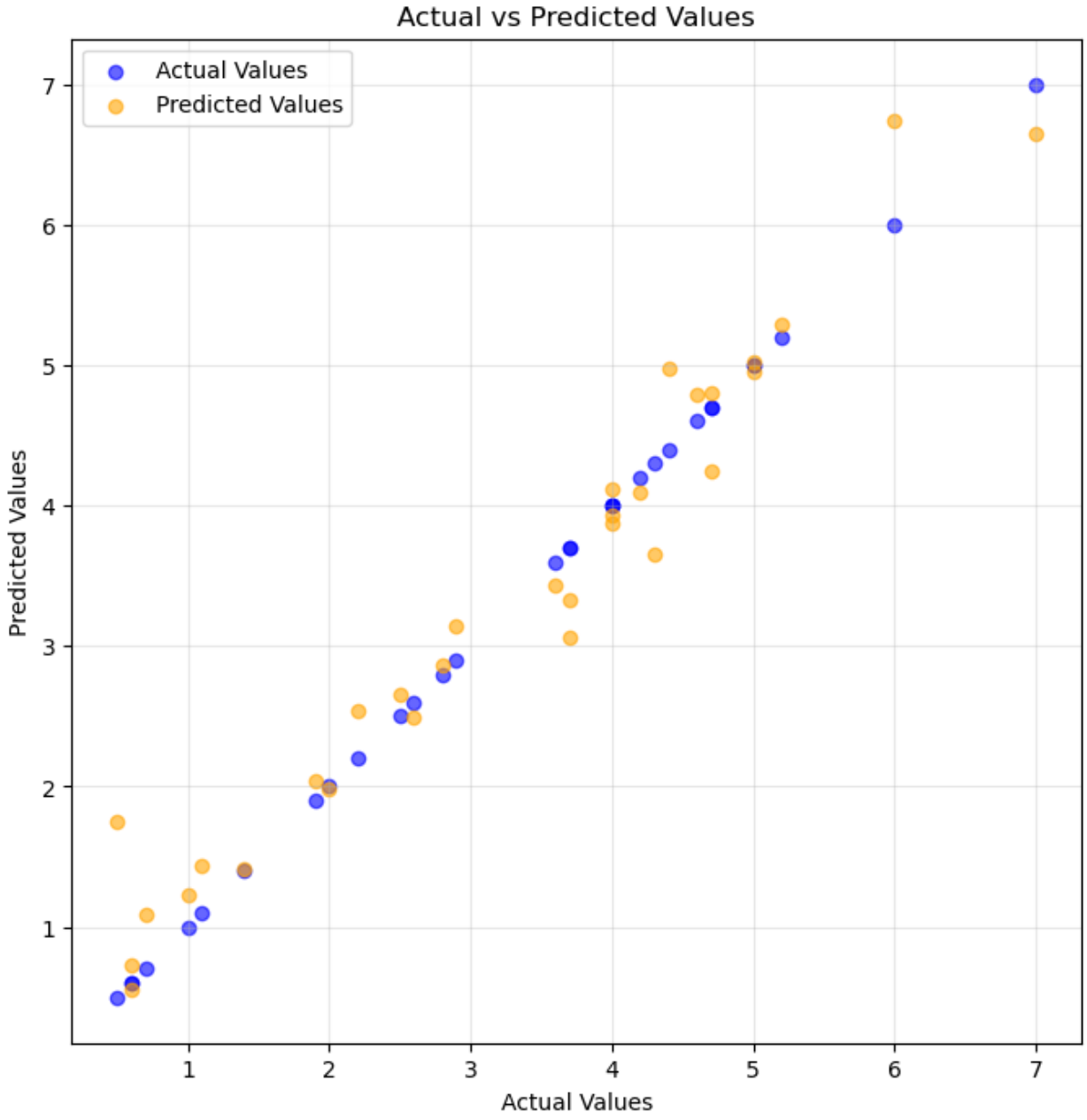
Helps to Identify **correlations** or **patterns** between pairs of features.



7. Confusion Matrix:



11. Random Forest Regression:



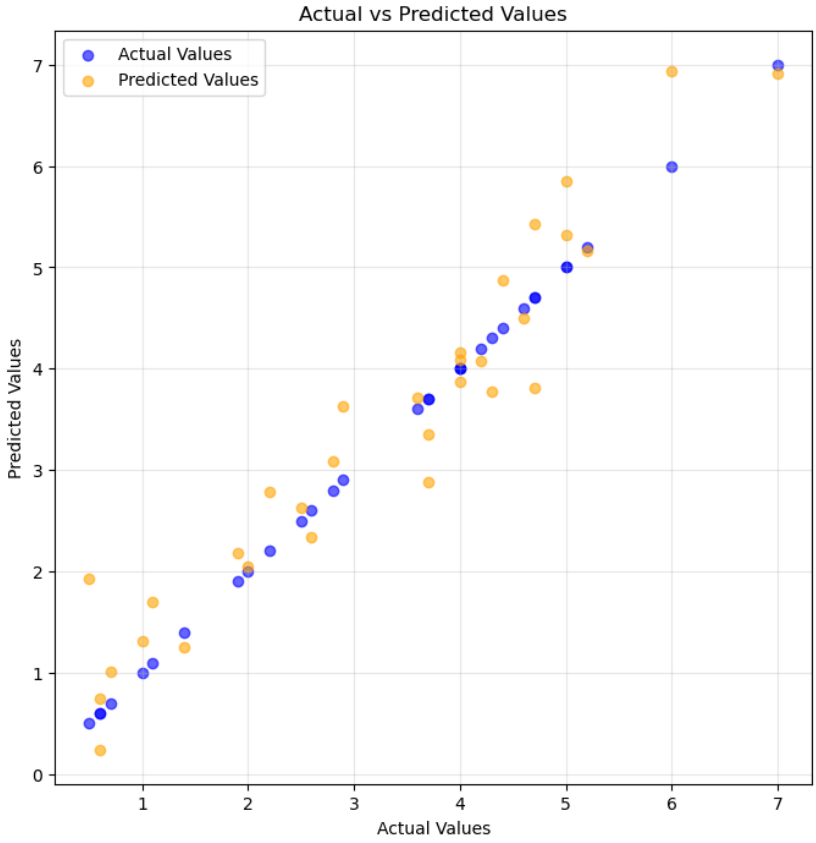

```
MAE: 0.2667951612903227
MSE: 0.1432184942741933
RMSE: 0.37844219409864077
R²: 0.9507440402408169
Adjusted R²: 0.9384300503010212
```

Testing with Own Dataset

```
Model: Random Forest Regressor
   efficiency         Predicted Efficiency
0       0.097               0.25535
1       0.377               0.35091
2       0.641               0.61893
3       0.419               0.60038

Metrics:
Mean Absolute Error (MAE): 0.09697250000000007
Mean Squared Error (MSE): 0.01478529997500001
R² Score: 0.6049114515902756
```

## 12. XG Boost Regression:




```
MAE: 0.3984354337376933
MSE: 0.26734340337399987
RMSE: 0.5170526118046401
R²: 0.9080547803186505
Adjusted R²: 0.8850684753983131
```

Testing with Own Dataset

```
Model: XGBoost Regressor
   efficiency           Predicted Efficiency
0       0.097              0.097925
1       0.377              0.205905
2       0.641              0.641142
3       0.419              0.492766

Metrics:
Mean Absolute Error (MAE): 0.06148230314254761
Mean Squared Error (MSE): 0.00867899676481803
R² Score: 0.768082335883439
```

Deep Learning (ANN)

1. We have 6 input nodes for our deep learning model (morphology,dye,Jsc,Voc,FF,Passivation) and 1 output node (efficiency)

2. There are 3 different architectures that have been used and experimented with.

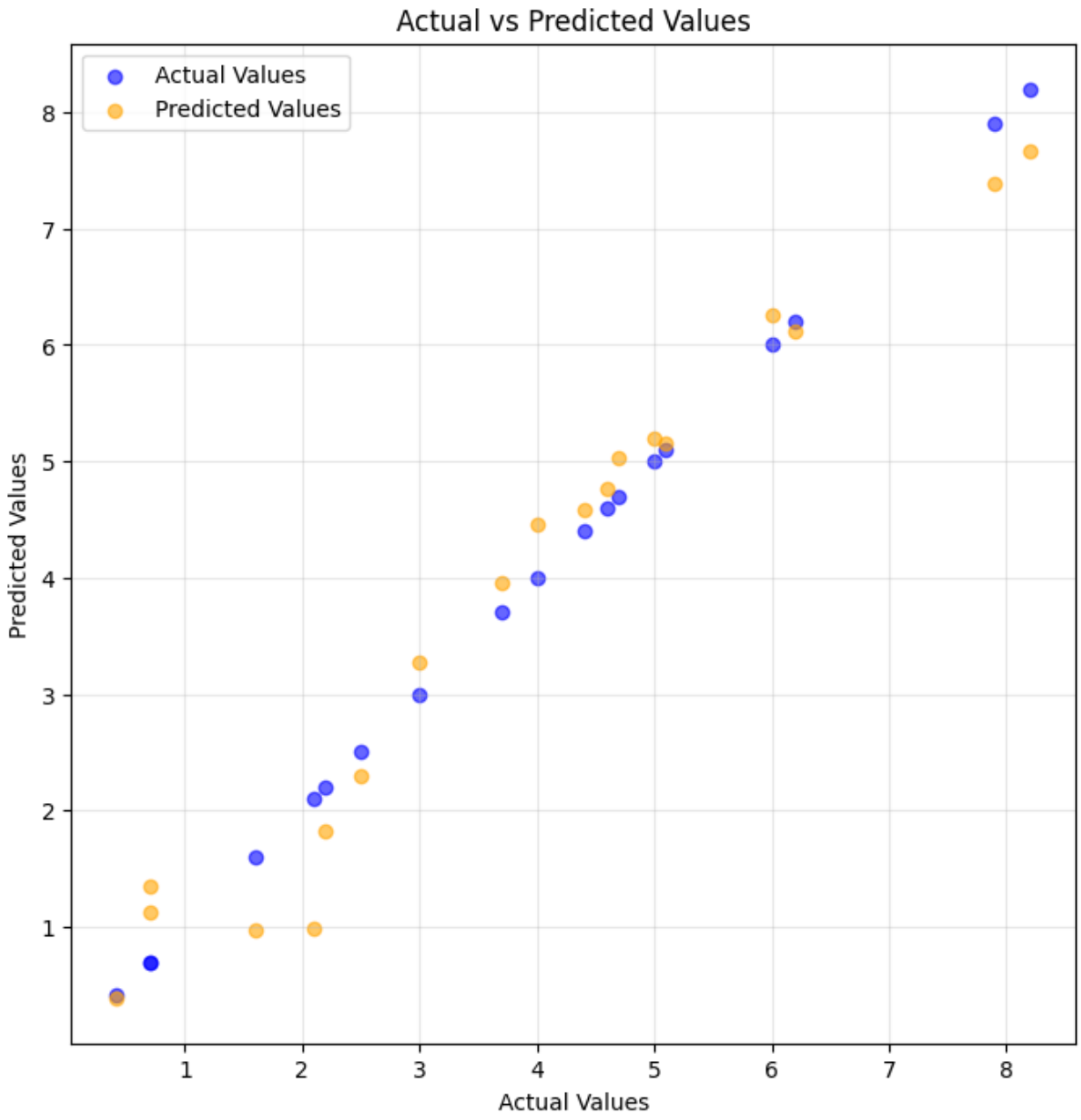
(A) Fully Connected Network

6 input nodes -> 32 Neuron(relu activation) ->16 Neuron(relu activation) -> 1 output node

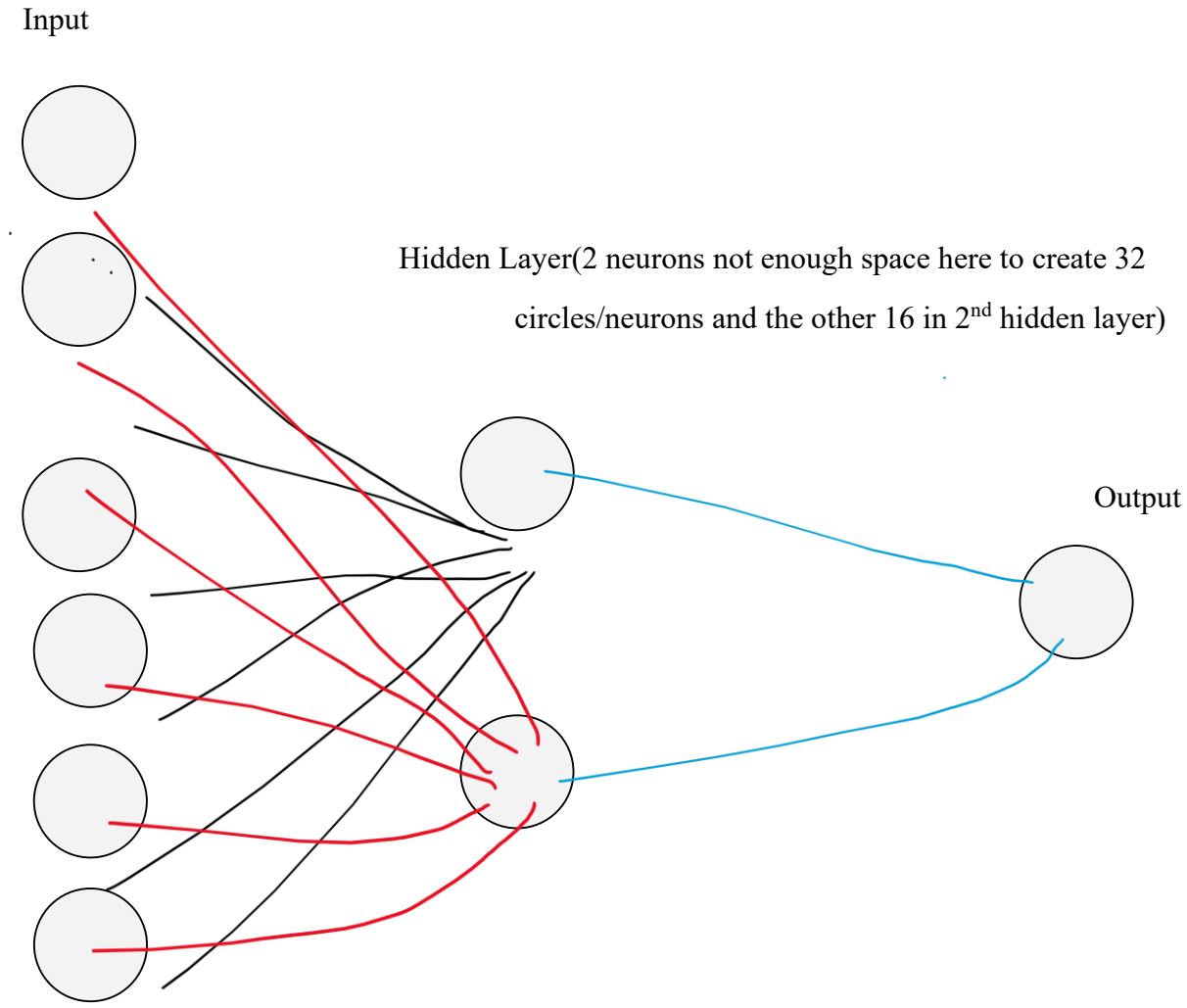
We used adam optimizer for training and MAE loss function, it was trained for 100,1000,2000 epochs respectively. Batch size of 1 was used and validation split was 30%

Dropout was used after hidden layer 1(20%) and same after hidden layer 2.

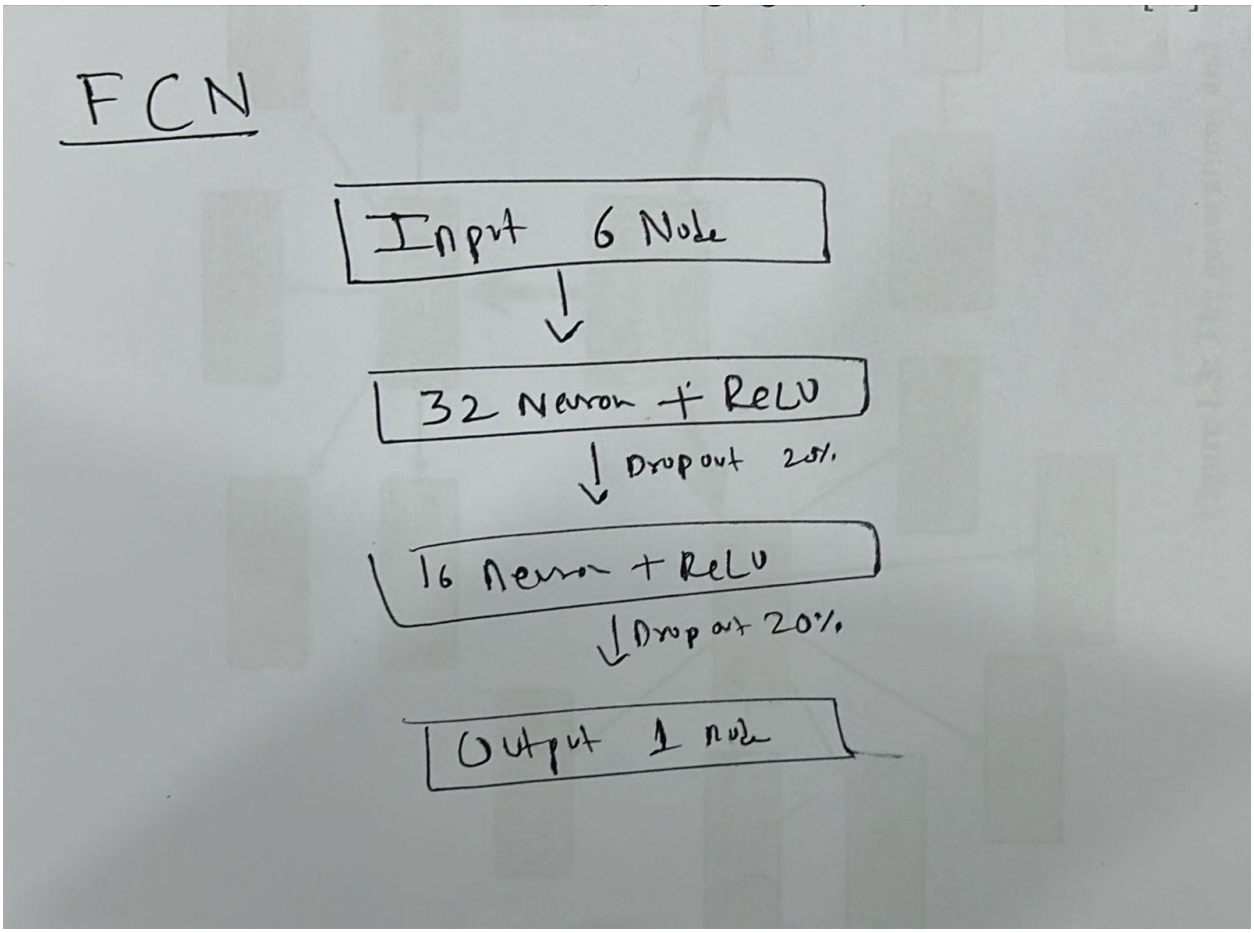
Metrics printed for test results are MSE, RMSE, MAE, R^2 (find in table 1)


Actual vs Predicted Values
Actual Values
Predicted Values
Predicted Values
Actual Values
1
2
3
4
5
6
7
8

@2000 epoch


Input
Hidden Layer(2 neurons not enough space here to create 32 circles/neurons and the other 16 in 2nd hidden layer)
Output

# FCN


Input 6 Node
32 Neuron + ReLU
Dropout 20%
16 Neuron + ReLU
Dropout 20%
Output 1 node

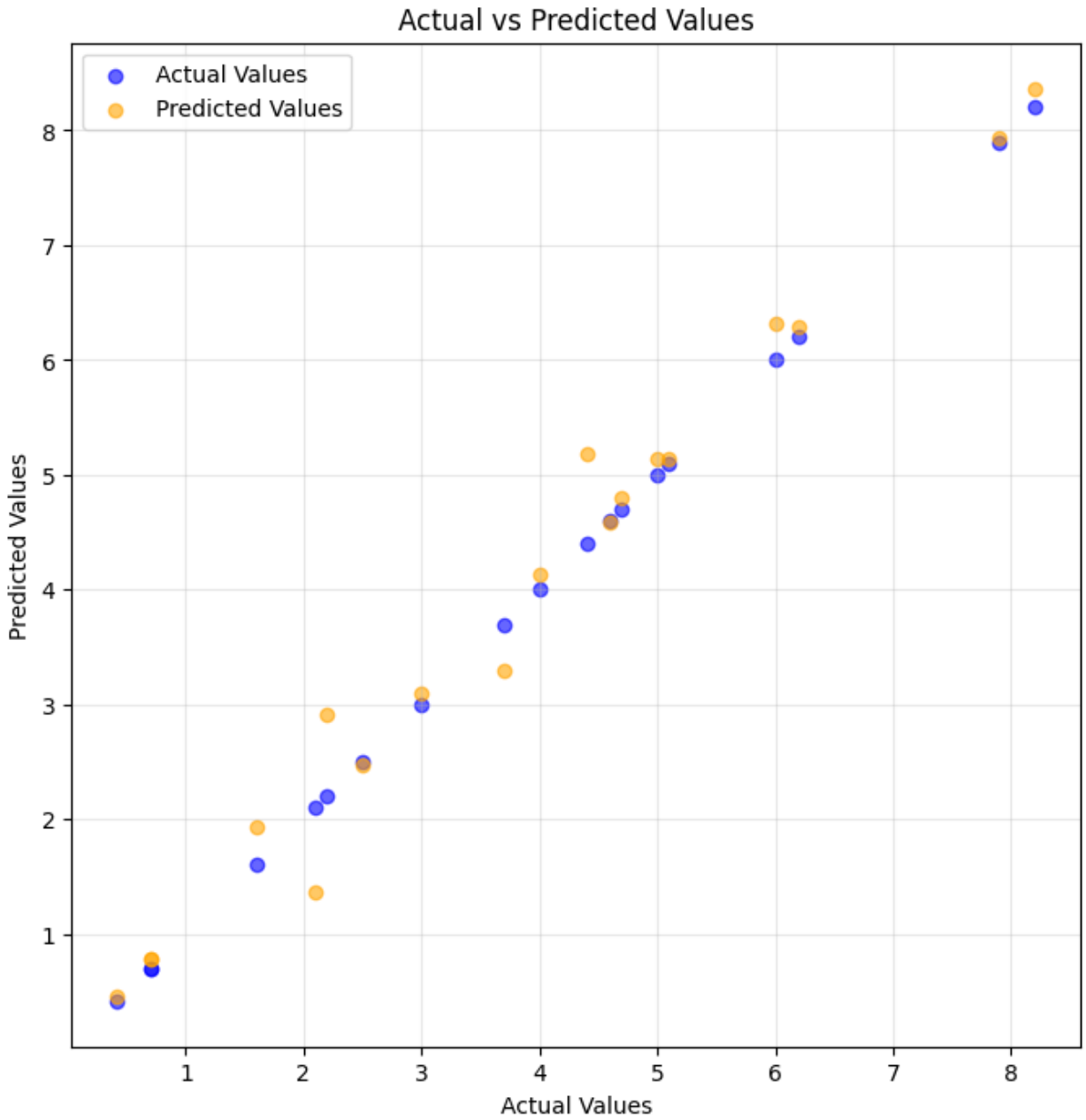
(B) Wide & Deep Neural Network

6 input nodes -|> deep branch (32 neurons with ReLU activation) -> deep branch(16 neurons+ReLU activation)

|-> wide branch (8 neurons with ReLU activation)

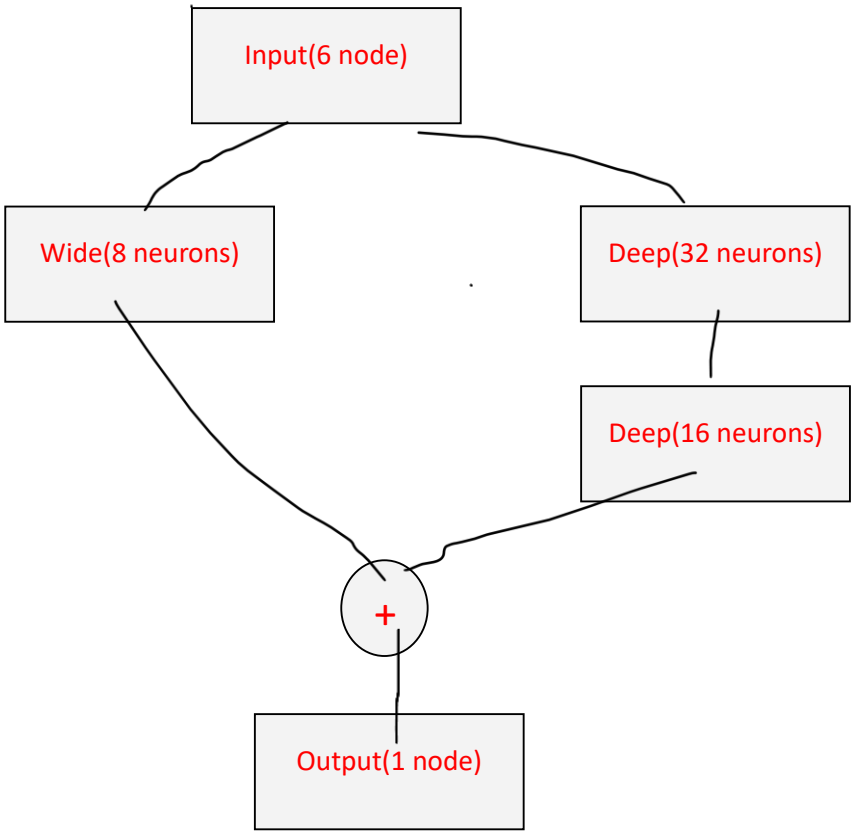
We used adam optimizer for training and MAE loss function, it was trained for 100,1000,2000 epochs respectively. Batch size of 1 was used and validation split was 30%

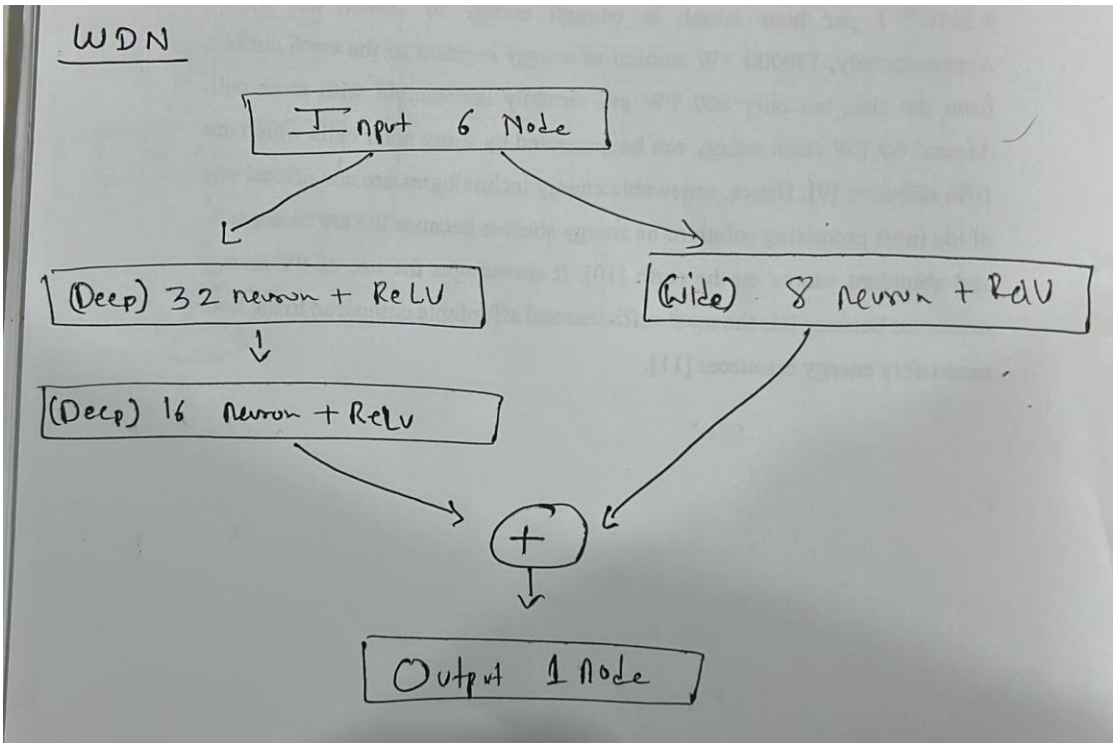
(find in table 1)




@2000 epoch

Concatenate deep(16 neuron)and wide(8 neuron) and connect to output node (1 node)

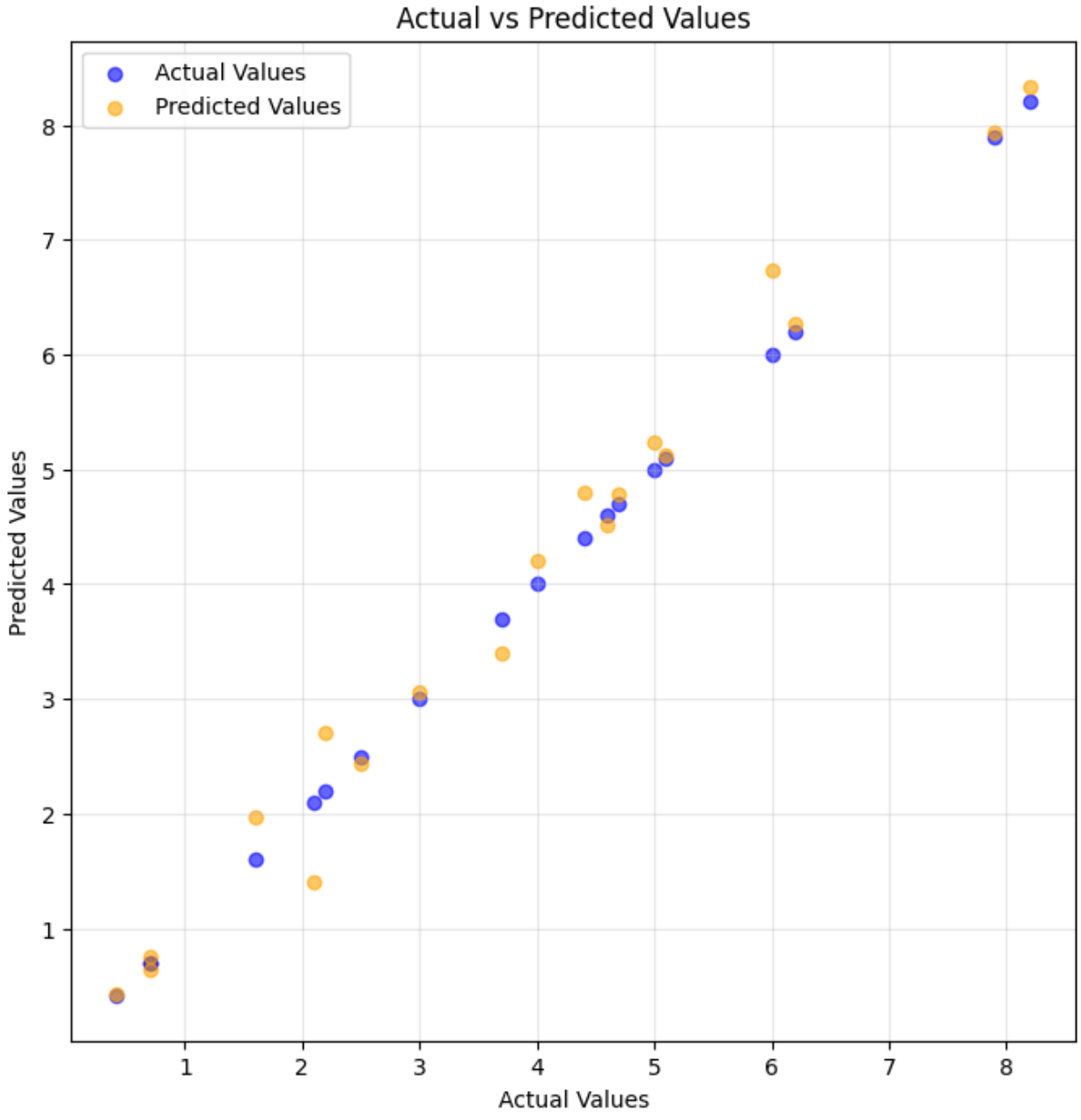


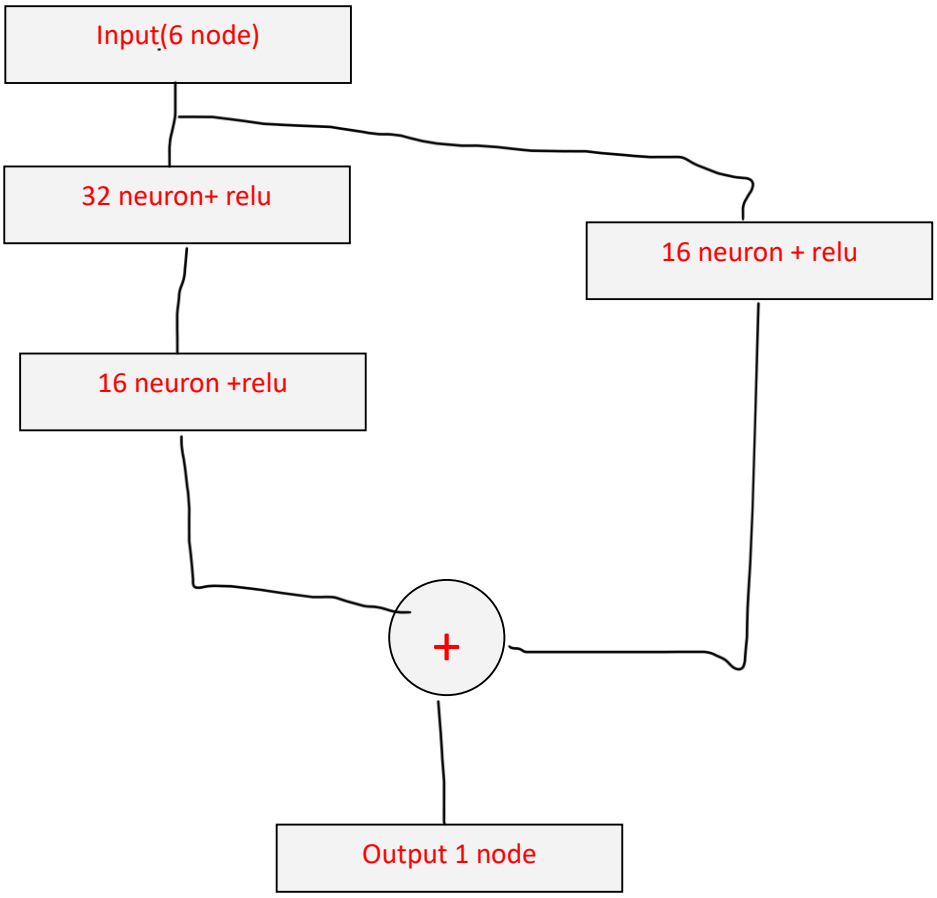
ResNet

We used adam optimizer for training and MAE loss function, it was trained for 100,1000,2000 epochs respectively. Batch size of 1 was used and validation split was 30%. (find in table 1)

Take input using 6nodes and pass one output 32 neuron and other one as a skip connection to 16 neurons, after the 32 forward it to 16 and then concatenate this 16 with the first input 16$^{th}$ neurons and pass to output node




@2000 epoch


Input(6 node)
32 neuron+ relu
16 neuron + relu
16 neuron +relu
+
Output 1 node

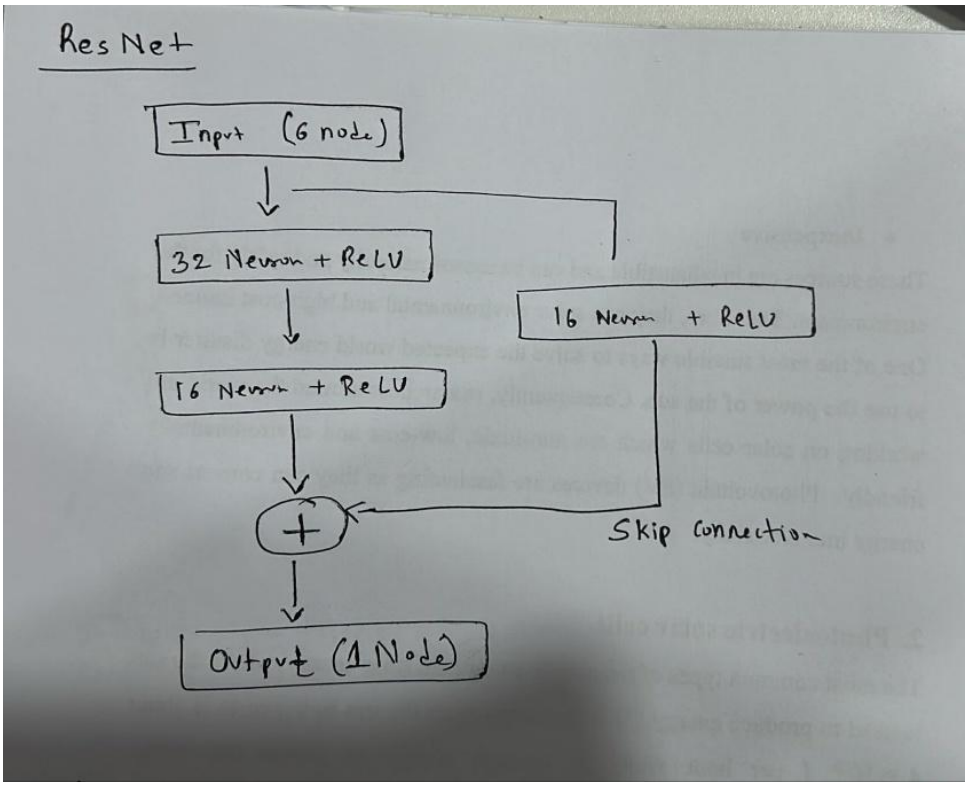

Res Net
Input (6 node)
32 Neuron + ReLU
16 Neuron + ReLU
16 Neuron + ReLU
+
Skip connection
Output (1 Node)

Table:1

# RESULTS

| ANN Model | Epoch | $R^2$ | MSE | MAE | RMSE |
|---|---|---|---|---|---|
| FCN | 100 | 0.95568 | 0.175146 | 0.35551 | 0.41850 |
| | 1000 | 0.97389 | 0.130427 | 0.29442 | 0.36128 |
| | 2000 | 0.97465 | 0.126727 | 0.30979 | 0.35598 |
| WDN | 100 | 0.98507 | 0.05898 | 0.14066 | 0.24287 |
| | 1000 | 0.98975 | 0.05124 | 0.15960 | 0.22637 |
| | 2000 | 0.97742 | 0.11288 | 0.22695 | 0.33598 |
| RNN | 100 | 0.98412 | 0.06273 | 0.18436 | 0.25047 |
| | 1000 | 0.98748 | 0.06255 | 0.17968 | 0.25010 |
| | 2000 | 0.98070 | 0.09648 | 0.21878 | 0.31062 |

Data of our own work

**100 epoch**

Wide and deep

Mean Squared Error (MSE): 0.0003164497629508092
$R^2$ Score: 0.9915439201301132
Actual vs Predicted Values:
Actual: 0.097, Predicted: 0.08108794689178467
Actual: 0.377, Predicted: 0.38698825240135193
Actual: 0.641, Predicted: 0.6572178602218628
Actual: 0.419, Predicted: 0.3935084044933319

ResNet

Mean Squared Error (MSE): 0.0005447617742089143
$R^2$ Score: 0.9854430320003497
Actual vs Predicted Values:
Actual: 0.097, Predicted: 0.11939069628715515
Actual: 0.377, Predicted: 0.34834563732147217
Actual: 0.641, Predicted: 0.6173132061958313
Actual: 0.419, Predicted: 0.4361920654773712

FCN

Mean Squared Error (MSE): 0.03197486208307774 6
$R^2$ Score: 0.1455768995309603
Actual vs Predicted Values:
Actual: 0.097, Predicted: 0.423934280872344 97
Actual: 0.377, Predicted: 0.4858217239379883
Actual: 0.641, Predicted: 0.6878960132598877
Actual: 0.419, Predicted: 0.5024986267089844

**1000 epoch**

Wide and Deep
Mean Squared Error (MSE): 0.0021335593660452654
$R^2$ Score: 0.9429876381066259
Actual vs Predicted Values:
Actual: 0.097, Predicted: 0.04678267240524292
Actual: 0.377, Predicted: 0.3225633502006531
Actual: 0.641, Predicted: 0.6507893800735474
Actual: 0.419, Predicted: 0.3646559417247772

ResNet

Mean Squared Error (MSE): 0.0035707233949141765
$R^2$ Score: 0.9045841528237722
Actual vs Predicted Values:
Actual: 0.097, Predicted: 0.14516222476959229
Actual: 0.377, Predicted: 0.36071228981018066
Actual: 0.641, Predicted: 0.6442612409591675
Actual: 0.419, Predicted: 0.31089186668395996

FCN

Mean Squared Error (MSE): 0.24828565984249865
$R^2$ Score: -5.634618242713288
Actual vs Predicted Values:
Actual: 0.097, Predicted: 0.7092263698577881
Actual: 0.377, Predicted: 0.8568938970565796
Actual: 0.641, Predicted: 1.1071228981018066
Actual: 0.419, Predicted: 0.8322224617004395

**2000 epoch**

Wide and Deep

Mean Squared Error (MSE): 0.0004987977
$R^2$ Score: 0.986671
Actual vs Predicted Values:
Actual: 0.097, Predicted: 0.111105
Actual: 0.377, Predicted: 0.381355
Actual: 0.641, Predicted: 0.666370
Actual: 0.419, Predicted: 0.45266

ResNet

Mean Squared Error (MSE): 0.000098338231
$R^2$ Score: 0.997372207
Actual vs Predicted Values:
Actual: 0.097, Predicted: 0.107581
Actual: 0.377, Predicted: 0.38455
Actual: 0.641, Predicted: 0.643644
Actual: 0.419, Predicted: 0.4337

FCN

Mean Squared Error (MSE): 0.0444087
$R^2$ Score: -0.18667
Actual vs Predicted Values:
Actual: 0.097, Predicted: 0.403018
Actual: 0.377, Predicted: 0.51551008
Actual: 0.641, Predicted: 0.761286
Actual: 0.419, Predicted: 0.6433515