The initial architecture uses input layer, two hidden layer and one output layer. The following table describe the comparison process between the initial architecture and modification architecture by adding more layers.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | layers | Total number of parameters | training time | performance | | | | | |
|  | RMSE | MAE | | MAE % | R2 score |
| Initial | Input layer  Hidden layer 1 (32)  Hidden layer 2 (16)  Output layer | 2199 | 138.27s | Training | 86838.88 | 65281.52 | | 10.87 | 0.82 |
| Testing | 86675.11 | 65106.03 | | 10.86 | 0.82 |
| Adding one layer | Input layer  Hidden layer 1 (32)  Hidden layer 2 (16)  Hidden layer 3 (8)  Output layer | 2337 | 214.18s | Training | 68291.39 | 49915.36 | | 8.16 | 0.89 |
| Testing | 68111.98 | 498636.92 | | 8.17 | 0.89 |
| Adding two layers | Input layer  Hidden layer 1(64)  Hidden layer 2 (32)  Hidden layer 3 (16)  Hidden layer 4 (8)  Output layer | 6081 | 141.50 s | Training | 60856.01 | 43779.58 | 7.08 | | 0.91 |
| Testing | 60662.88 | 43693.85 | 7.08 | | 0.91 |

According to the comparison results, the model that includes more hidden layers can slightly improve performance. However, if the architecture includes too many hidden layers, it may take significantly more time to train. Nevertheless, deeper architectures are capable of learning more complex patterns.

For larger dataset, adding the hidden layer increase the model performance because of learning the complex pattern using many dataset without overfitting. For small datasets, adding more layers can easily lead to overfitting, as the model gains high capacity with many parameters but lacks sufficient data to learn generalizable patterns.