

1. Hyper Parameters -

- a. Solver
- b. Initial learning rate, α
- c. Adjustment of learning rate
- d. Number of hidden layers and neurons
- e. Activation function
- f. Regularization parameter L2

2. In this paper, authors use a publicly available dataset, named, “COVID-19 Data Repository” which contains the data for the coronavirus patients describing the number of patients in a certain location, for each day from 22nd of January 2020 to 12th of March 2020. The repository operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) and supported by ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab. The repository is available at <https://github.com/CSSEGISandData/COVID-19>

3. K-fold Cross Validation provides a solution by dividing the data into folds and ensuring that each fold is used as a testing set at some point. The data set is divided into k subsets, and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set and the other k-1 subsets are put together to form a training set. Then the average error across all k trials is computed. The advantage of this method is that it matters less how the data gets divided.

In the paper, the dataset is split into k subsets where k=5 for implementation. Then, each of them is used as a testing set and the remaining k – 1 subsets are used as a training dataset. Then the authors presented the outcome as the average of achieved scores, with standard deviation noted. The outcome has been constructed using scikit-learn library of Python 3.8 .

4. The Grid Search algorithm basically tries all possible combinations of parameter values and returns the combination with the highest accuracy. To determine the best hyperparameter combination, authors used the grid search algorithm. Each of the combinations is used to train the MLP. To avoid the possibility of poor solutions due to the initial random setting of the weights, each set of hyperparameters is used for training three times.

Four hidden layers have been used by authors in the MLP mentioned in the paper .

5. 1. It has the ability to learn faster than other activation functions which determines it allows the network to converge very quickly
2. It removes negative values from the model as it only generates positive values.