The project utilizes data from the divisions and states of Myanmar, comprising 19,513 entries and 2 features, for a classification task. Before the classification process, a data preprocessing stage is performed. This stage involves cleaning the data to address missing values, duplicates, and inconsistencies. Additionally, categorical variables are converted into numerical formats to make the data suitable for machine learning algorithms. After preprocessing, the dataset is split into two parts: a training set with 13,659 records and 13,003 features, and a testing set with 5,854 records and the same number of features.

For the model training process, the original implementation uses a neural network consisting of an input layer, two hidden layers, and one output layer. To improve performance, modifications were made to the model. First, one additional hidden layer was added, resulting in a three-hidden-layer architecture. Later, a second hidden layer was added, creating a network with four hidden layers in total. In the process, ReLU use as activation function and Softmax use as classification model. The results of each configuration—original and modified—are compared and summarized in the following table, highlighting performance differences between the various architectures.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | layers | Total number of parameters | training time | performance | |
|  | ROC AUC score |
| Initial | Input layer  Hidden layer 1 (32)  Hidden layer 2 (16)  Output layer | 416,962 | 213.37 seconds | Training | 0.9970 |
| Testing | 0.6967 |
| Adding one layer | Input layer  Hidden layer 1 (32)  Hidden layer 2 (16)  Hidden layer 3 (8)  Output layer | 416,954 | 220.37 seconds | Training | 0.9956 |
| Testing | 0.6701 |
| Adding two layers | Input layer  Hidden layer 1(64)  Hidden layer 2 (32)  Hidden layer 3 (16)  Hidden layer 4 (8)  Output layer | 835,162 | 336.57 seconds | Training | 0.9950 |
| Testing | 0.6735 |

Accordingly, to the results, the original model had the best test ROC AUC. Adding layers increased parameters and training time but slightly reduced test performance. Therefore, I think in this process deeper models cannot improve classification effectiveness. And then, overfitting occurs when a model performs well on training data but poorly on unseen test data.