Data Science Project Pt. 2 (MAST90107)

Final Report, Semester 2

Kartika Waluyo, 1000555 Vrinda Rajendar Rajanahally, 1129446

Contents

Table 1: XGBoost Final Model Result

	MAE	MSE
Training	0.1775177	0.0539348
Test	0.6653771	0.7352190

Table 2: Neural Network Final Model Result

	MAE	MSE
Training	0.5993975	0.6128888
Test	0.6464047	0.6931747

$$F_n = F_{n-1} + \eta \Delta_n \quad (1)$$

$$F_m(X) = F_{m-1}(X) + a_m h_m(X, r_{m-1}) \quad (2)$$

$$ENSpred = \frac{1}{2} NNpred + \frac{1}{2} XGBpred = \frac{1}{2} \begin{bmatrix} \hat{y}_{11} & \cdots & \hat{y}_{1p} \\ \vdots & \ddots & \\ \hat{y}_{n1} & \cdots & \hat{y}_{np} \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \hat{z}_{11} & \cdots & \hat{z}_{1p} \\ \vdots & \ddots & \\ \hat{z}_{n1} & \cdots & \hat{z}_{np} \end{bmatrix} \quad (3)$$

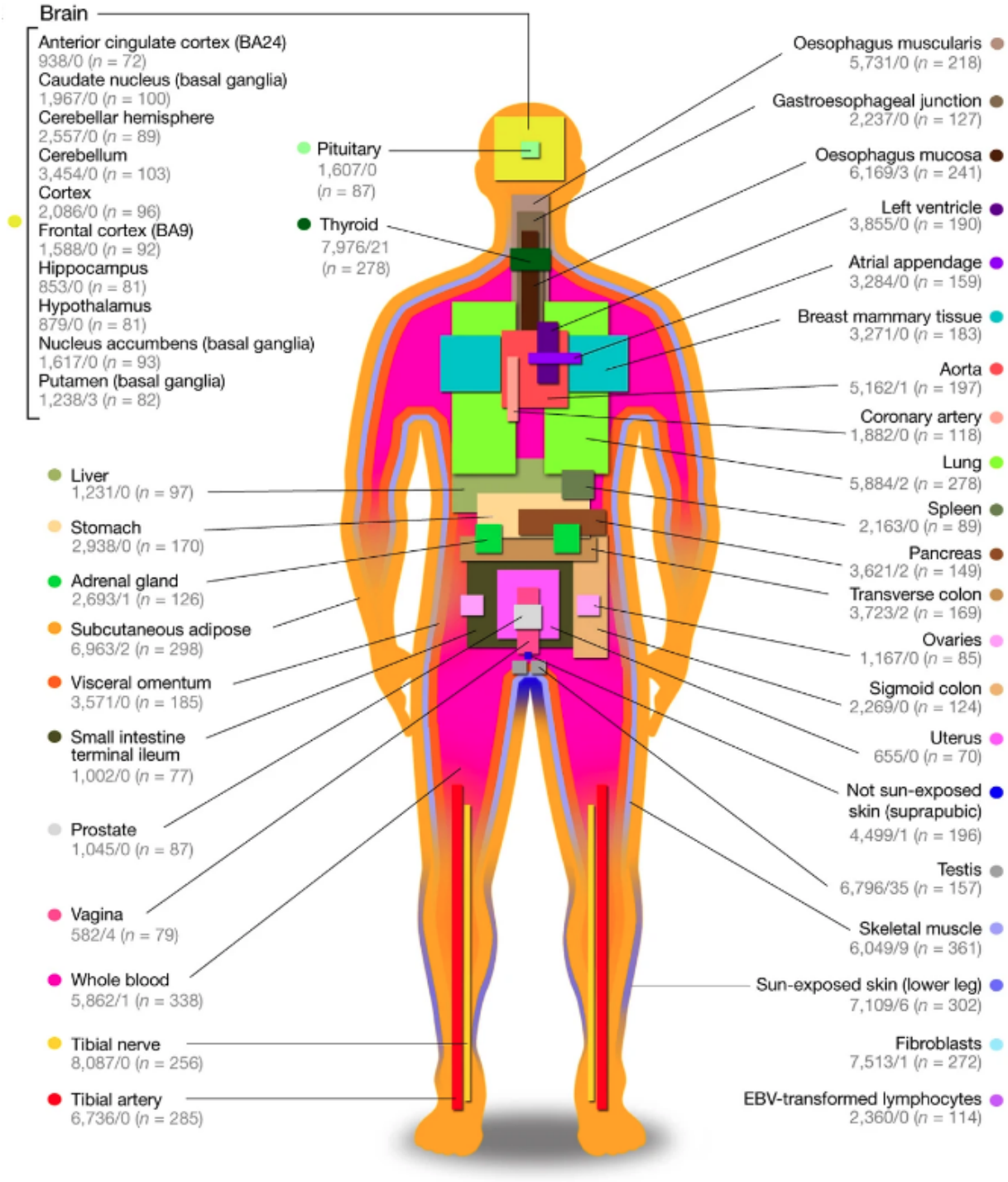

Figure 1: Sample from Various Tissues (GTEx Portal, 2021)

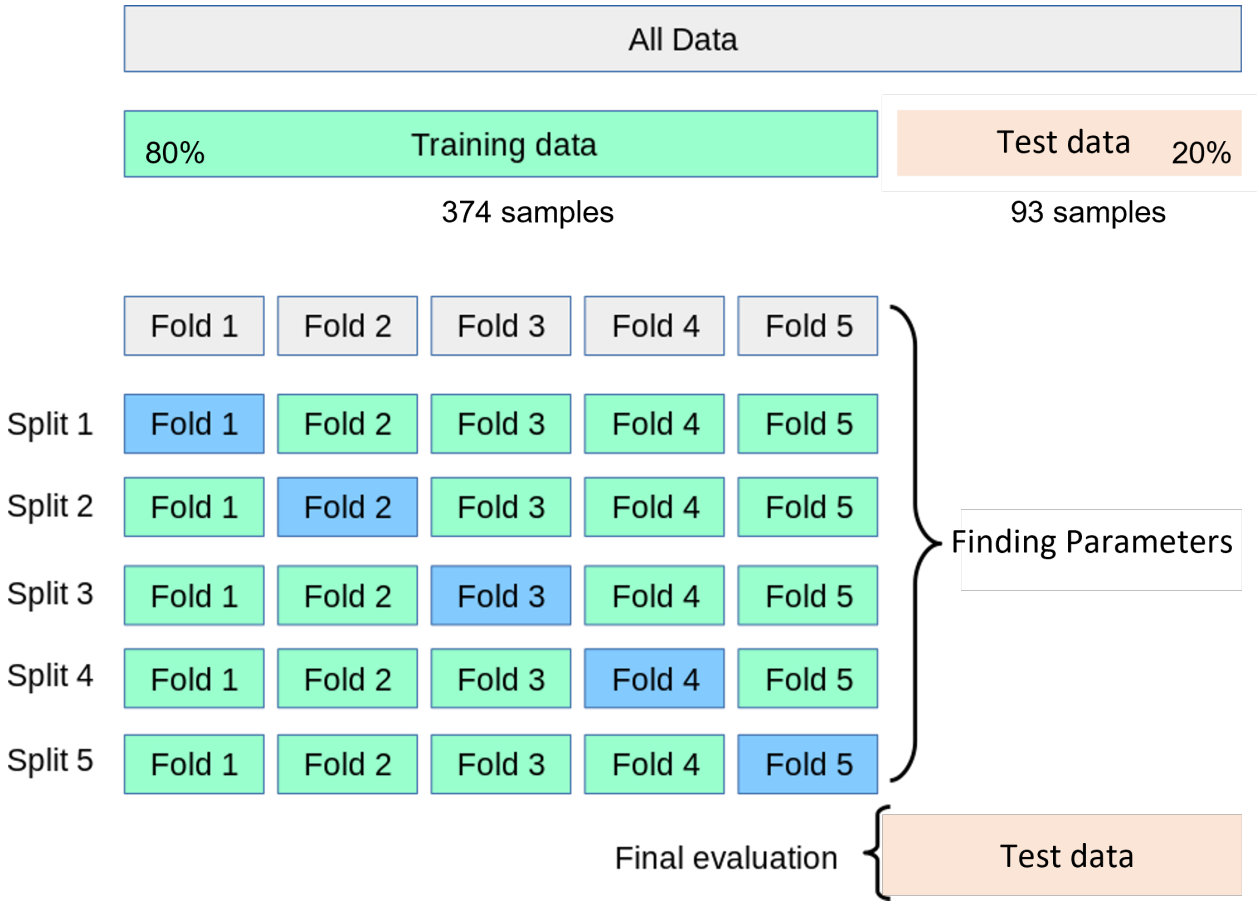

Figure 2: Data Splitting for Training, validation, and Test Set (3.1. Cross-Validation: Evaluating Estimator Performance, 2021)

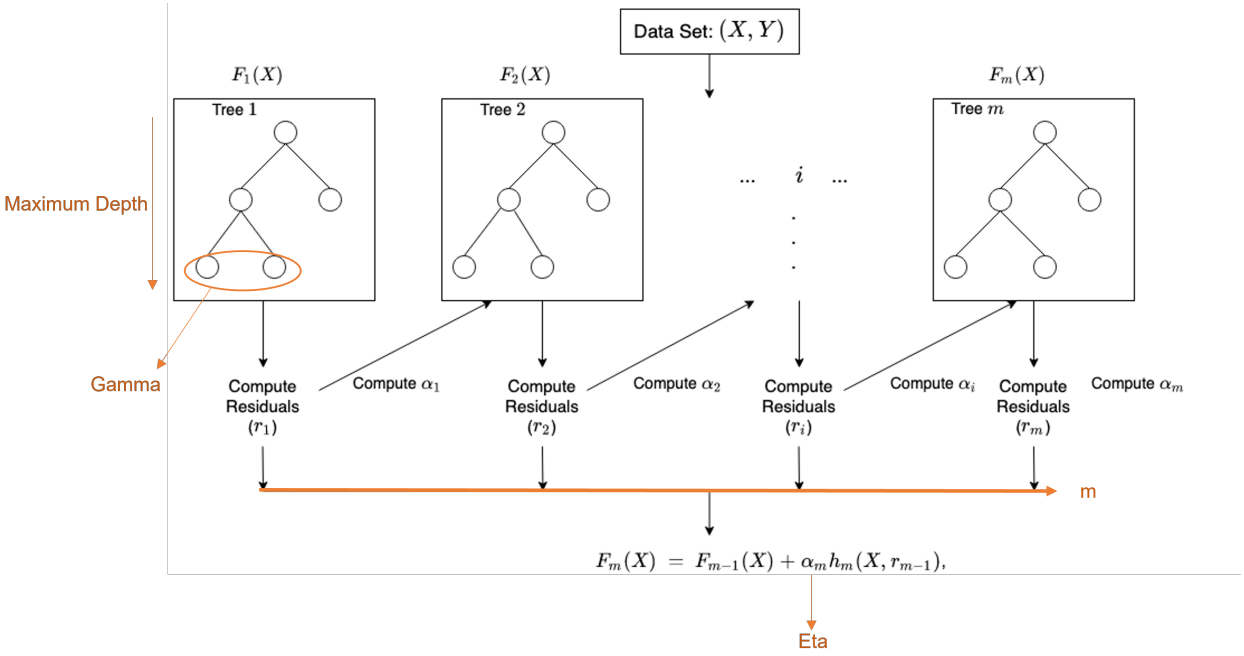

Figure 3: Rough Illustration on How XGBoost Works (How XGBoost Works - Amazon SageMaker, 2021)

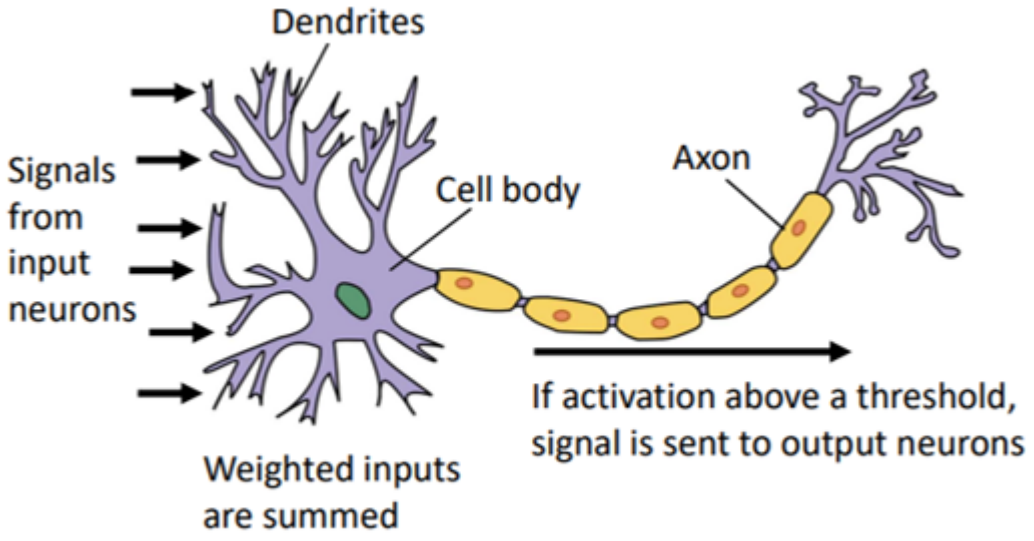

Figure 4: Biological Neural Network (Welcome to SEER Training | SEER Training, 2021)

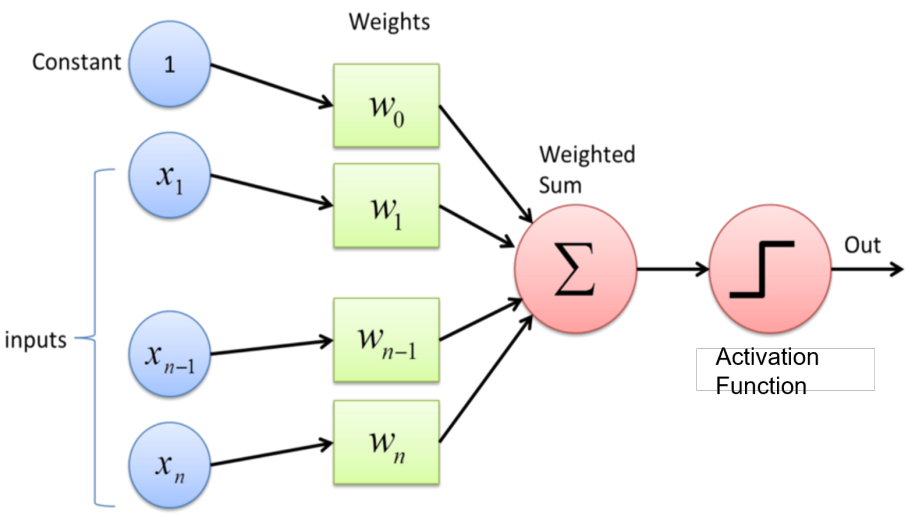

Figure 5: Perceptron (Sand et al., 2019)

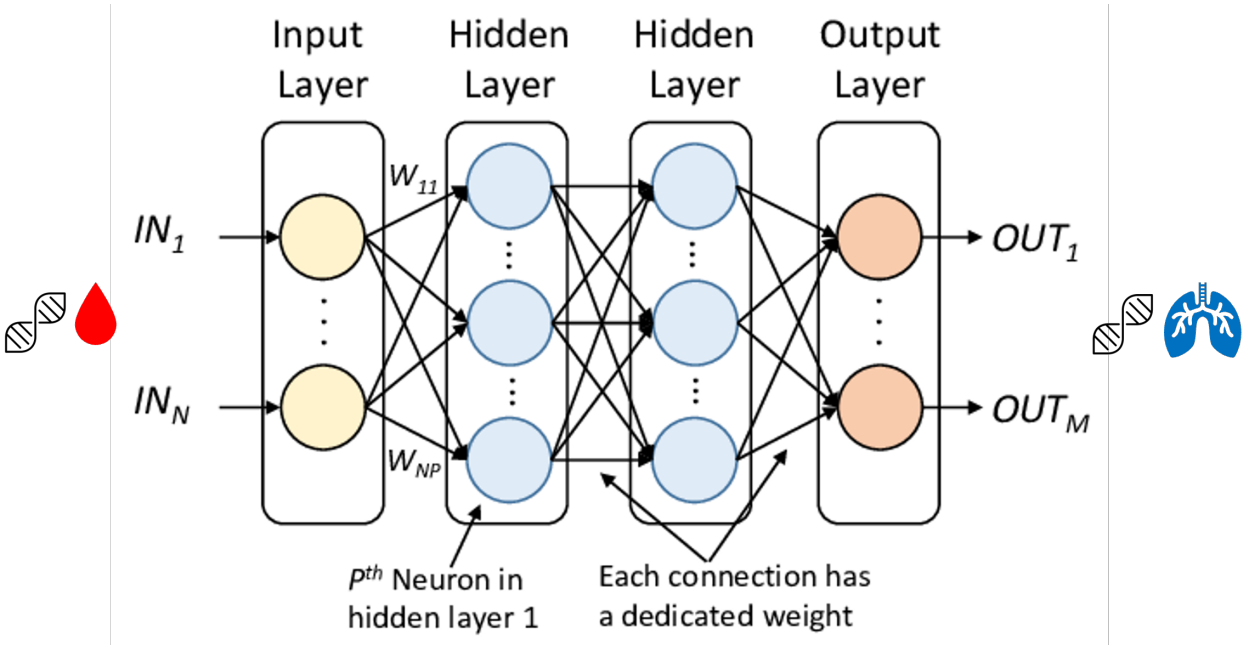

Figure 6: Multi-layer Neural Network (An Overview on Multilayer Perceptron (MLP), 2021)

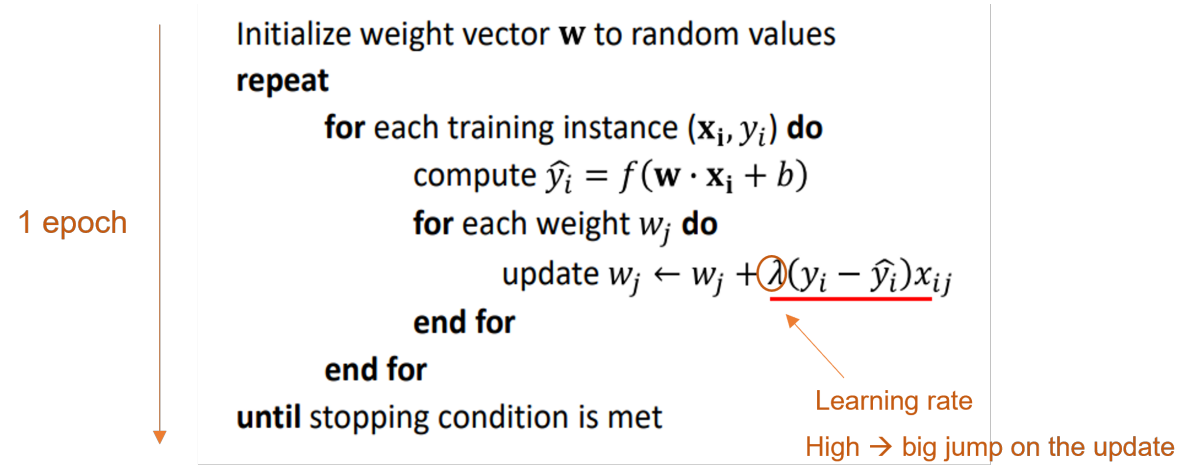

Figure 7: Neural Network Algorithm

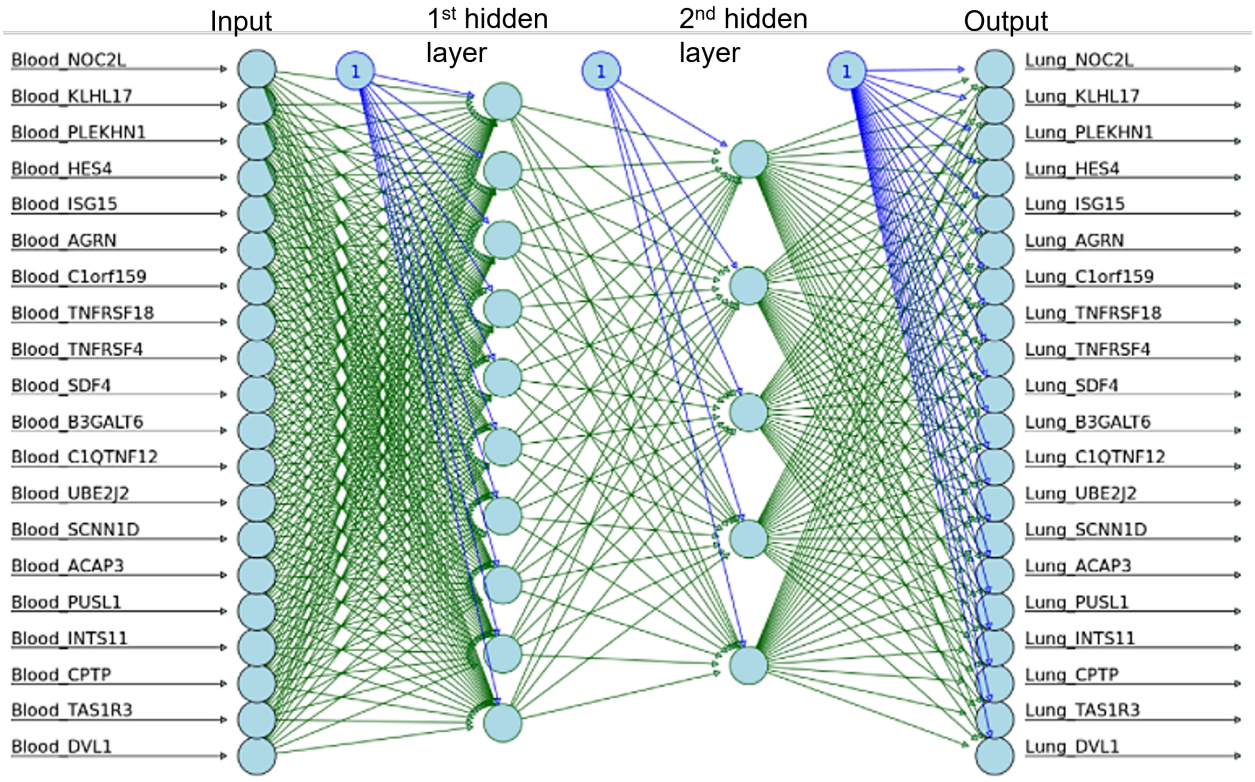

Figure 8: Illustration of Our Neural Network

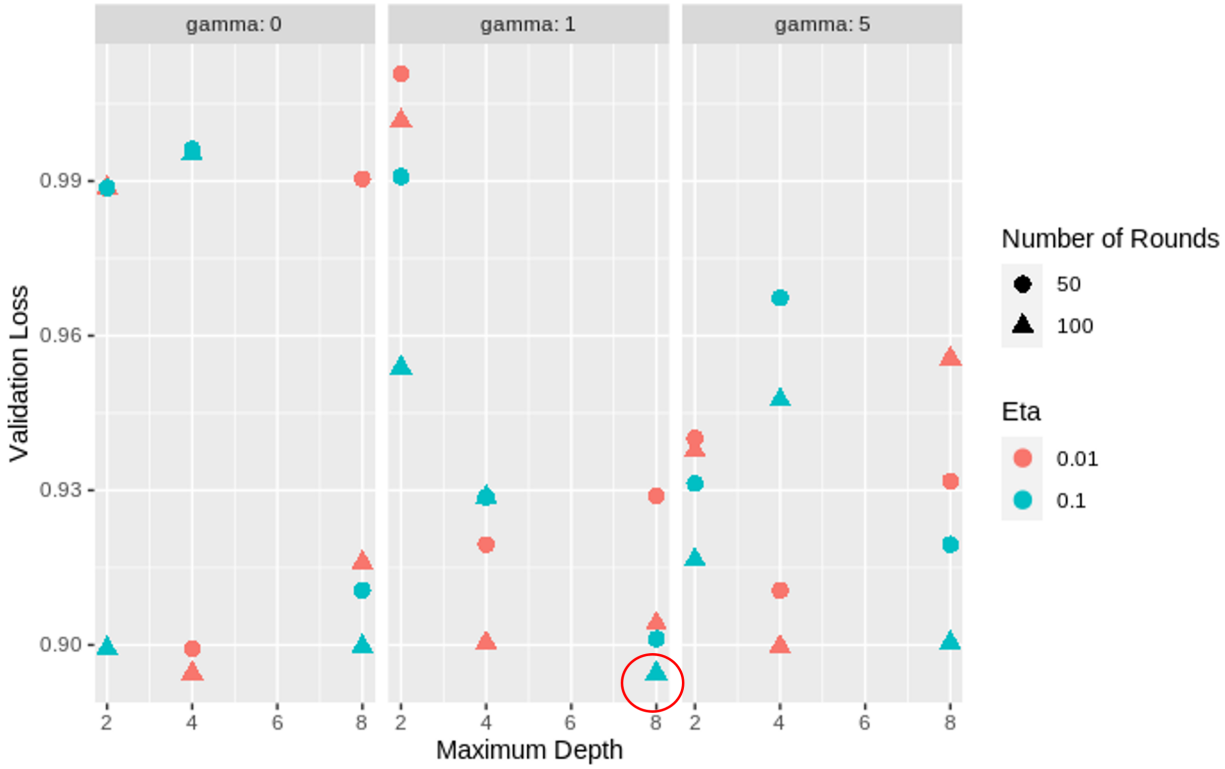

Figure 9: XGBoost Gridsearch Results

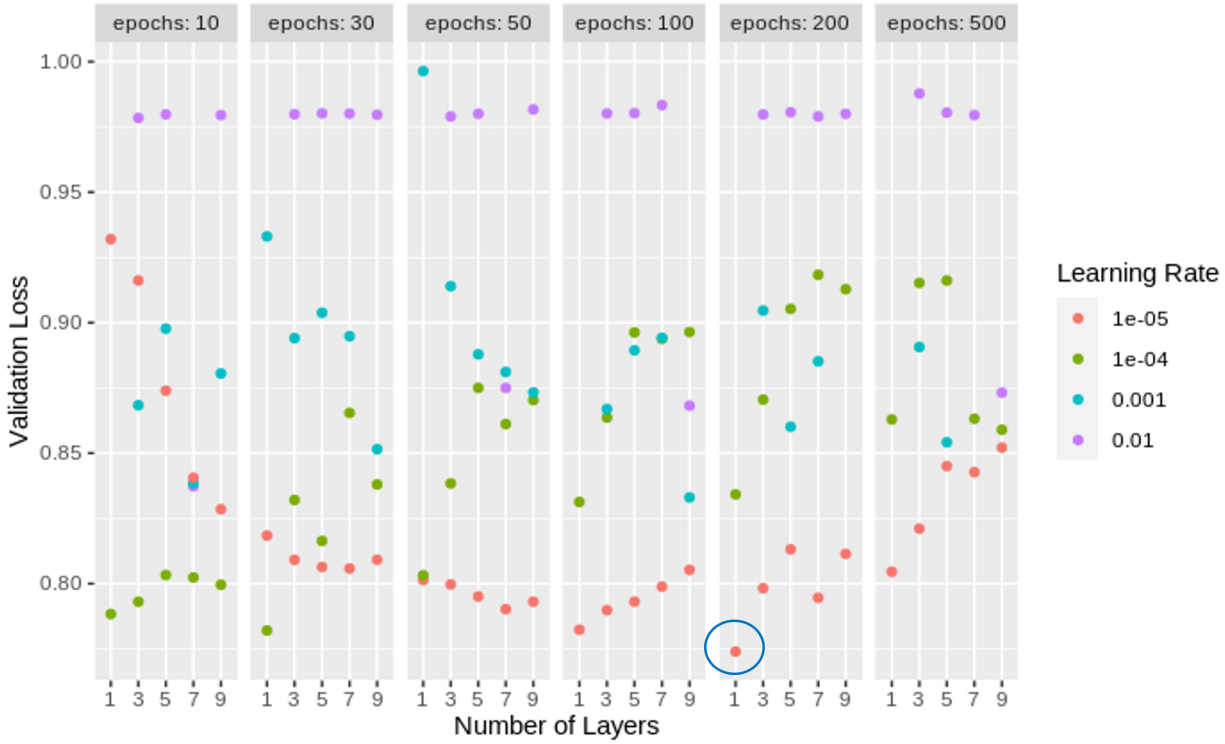

Figure 10: Neural network Gridsearch Results

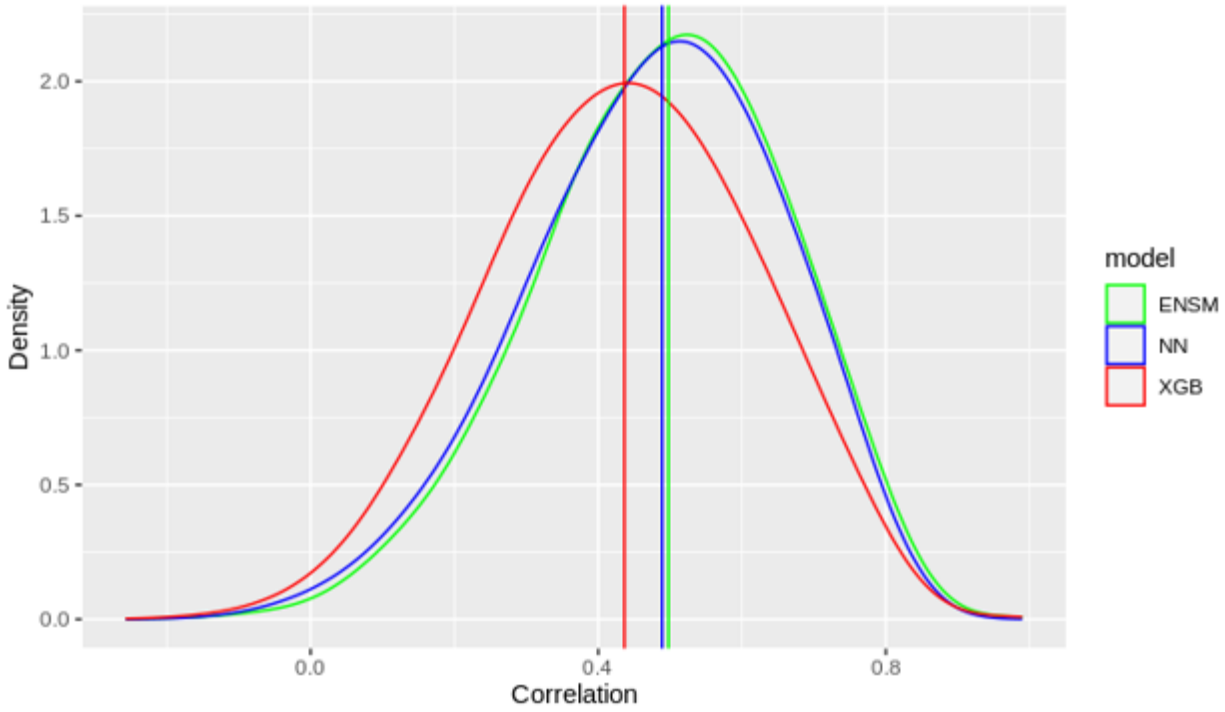

Figure 11: Test Prediction Correlation with Median Line

Table 3: Median Correlation

	Training	Test
XGBoost	0.9909750	0.4364947
Neural Network	0.6278658	0.4887672
Ensemble	0.9205496	0.4976427