| # of Nodes | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 6,000 | 7,000 | 8,000 | 9,000 | 10,000 |
|---|---|---|---|---|---|---|---|---|---|---|
| Execution Time | 1.905 | 3.209 | 4.850 | 6.272 | 8.102 | 9.826 | 11.371 | 13.248 | 15.462 | 17.084 |

Table 1: The efficiency analysis of the GTMP model, measured in seconds, examines the relationship between the number of nodes in the trees, ranging from 1,000 to 10,000, and the execution time. The latter reflects the training time for one epoch of 1,000 trees, averaged over 100 runs. The Pearson correlation coefficient between the number of nodes and the execution time is 0.999, and the p-value is $1.76 \times 10^{-11}$. These results indicate a strong linear time complexity of our proposed method with respect to the number of nodes, demonstrating its efficiency and scalability.

| | Neuron - Classification ($\uparrow$) | | | | | River - Regression ($\downarrow$) | | |
|---|---|---|---|---|---|---|---|---|
| | lps-glia | lps-inter | lps-pc | 5xfad-glia | 5xfad-pc | $\mu$ | $D$ | $r$ |
| $\{\alpha\}_1$ | 0.9836 | 0.9996 | 0.9872 | 0.9011 | 0.9992 | 0.0041 | 76.7699 | 33.0287 |
| $\{\alpha\}_2$ | 0.9779 | 0.9999 | 0.9764 | 0.8890 | 0.9985 | 0.0038 | 77.3823 | 32.0424 |
| lr=$1 \times 10^{-3}$ | 0.9836 | 0.9996 | 0.9804 | 0.8984 | 0.9991 | 0.0042 | 76.7699 | 33.0287 |
| lr=$5 \times 10^{-4}$ | 0.9810 | 0.9958 | 0.9872 | 0.9011 | 0.9992 | 0.0041 | 76.9832 | 33.8315 |
| bs=16 | 0.9836 | 0.9996 | 0.9804 | 0.9003 | 0.9991 | 0.0045 | 76.7699 | 33.0287 |
| bs=64 | 0.9849 | 0.9996 | 0.9751 | 0.9011 | 0.9991 | 0.0041 | 77.5353 | 34.1242 |

Table 2: Here is the parameter sensitivity analysis for the GTMP model, focusing on the impact of three types of hyperparameters across eight datasets. Initially, we explore the effect of the hyperparameter set $\{\alpha\} = \{\alpha_1, \alpha_2, \alpha_3\}$, which modulates the influence of each ancestor in Equation 4. The notation $\{\alpha\}_1$ signifies uniform weighting, achieved by setting each $\alpha$ value to 1, thereby distributing equal influence among ancestors. Conversely, $\{\alpha\}_2$ represents a decreasing weighting scheme, with $\alpha_1 = 1$ maintaining full influence, $\alpha_2 = 0.8$ indicating moderately reduced influence, and $\alpha_3 = 0.5$ depicting significantly diminished influence. Subsequently, we evaluate the model's sensitivity to the initial learning rate, comparing the effects of setting it to $1 \times 10^{-3}$ and $5 \times 10^{-4}$. Lastly, we assess the impact of selecting different batch sizes, as 16 and 64, on the model's performance.