

Abstract

RNA-binding proteins (RBPs) are a crucial class of molecules that regulate gene expression, RNA splicing, RNA transport, RNA stability, and translation. These processes are fundamental for normal cellular function and development. Alterations in RBP function can lead to a variety of diseases, including cancer, neurodegenerative diseases, and disorders of RNA metabolism. RBPs are thus important targets for therapeutic development. Their ability to interact with various RNA molecules makes them essential components of cellular processes. Despite their significance, the application of machine learning techniques to RBP classification remains in its early stages. This study investigates the impact of network size on the performance of machine learning models for RBP classification. The results demonstrate that the network size of 16 yields the lowest training losses and the highest training accuracy. However, the differences in performance across different network sizes are marginal, indicating that the network is not overly sensitive to size changes. The validation accuracy consistently remains around 0.5, suggesting that the network generalizes well to unseen data. The training time varies across network sizes, with the size of 16 taking the longest and the size of 64 taking the shortest. Overall, the findings suggest that machine learning can be effectively employed for RBP classification, and that network size plays a role in optimizing performance.

Keywords—RBP, Transposed Layer, Dense Layer