

CAITLIN CORCORAN, Department of Chemistry, Shepherd University, Shepherdstown, WV, 25443, and JASON MILLER, Department of Computer Science, Mathematics, and Engineering, Shepherd University, Shepherdstown, WV, 25443. Effect of Dropout on RNA Classification by CNN.

Long RNA sequences can be classified as long non-coding RNA (lncRNA) or protein-coding messenger RNA (mRNA). Automatic classification, based on sequence alone, could benefit biology and medical science. We trained and evaluated a convolutional neural network (CNN) to classify human RNA sequences. The CNN incorporated dropout, a technique that restricts the network to a random portion of its neurons during training. Dropout can reduce overfitting, which means relying on irrelevant aspects of the data to “memorize” the training set. We varied the dropout rate during training and measured the accuracy during testing for an RNA classification task. At dropout rates of 0.5, 0.6, 0.7, and 0.8, the CNN test accuracy was 93.15%, 90.50%, 89.95%, and 88.35%. We conclude that dropout rates above 50% did not improve learning. In the future, we hope to measure the effects of other hyperparameters and models for this classification task.