ANDREW MOSEMAN, CME, Shepherd University, Shepherdstown, WV, 25443, and SAMON BHATTA, CME, Shepherd University, Shepherdstown, WV, 25443, and JASON MILLER, CME, Shepherd University, Shepherdstown, WV, 25443. Machine Learning applied to an RNA Classification Problem.

Long RNA sequences can be classified as protein-coding messenger RNA (mRNA) or as long non-coding RNA (lncRNA). RNA classification based on sequence alone is a bioinformatics challenge with potential to enlarge our understanding of the causes of many human diseases. We chose to compare two classification approaches: (1) applications of correlation statistics and visualization tools, and (2) training a convolutional neural network. Both approaches used human RNA sequences from the public GenCode database. We found that the machine learning approach was superior, achieving an accuracy of 87.37%, which is not as high as some published classifiers. These results indicate that machine learning techniques are a more effective solution to the problem of human RNA classification compared to the other techniques tested. For future work, we propose to build, train, and compare other convolutional neural network models for this classification task.