ABSTRACT:
A growing number of applications rely on Neural Networks for making decisions. Yet, we still do not have a good understanding of how decisions are made and whether they are correct. Similarly, when using Neural Networks, it is not clear which structure should be used. It is not clear yet how the structure of a NN influences its performance. However, having a network with unnecessary nodes or connections will definitely contribute to its space and then operation efficiency. Specifically, in this work, we look at the problem of identifying edges and nodes in the network that we can remove without compromising its overall performance. This work is part of a larger effort about understanding the mechanics of neural networks, including via visualization, so that Data Scientists and Software Engineers (SWEs) who use Neural Networks can better understand them.

In this work, we study the importance (and effect of removal) of specific nodes in the network as well as the meaning of weight values as it relates to the importance of these nodes. We report our promising preliminary results and draw directions for future work.