A Novel Approach to Adopt Explainable Artificial Intelligence in X ray Image Classification

Shaw-Hwa Lo

Department of Statistics, Columbia University, U.S.A.

Abstract

Robust "Blackbox" algorithms such as Convolutional Neural Networks (CNNs) are known for making high prediction performance. However, the ability to explain and interpret these algorithms still require innovation in the understanding of influential and, more importantly, explainable features that directly or indirectly impact the performance of predictivity. In view of the above needs, this study proposes an interaction- based methodology – Influence Score (I score) – to screen out the noisy and non-informative variables in the images hence it nourishes an environment with explainable and interpretable features that are directly associated to feature predictivity. We apply the proposed method on a real-world application in Pneumonia Chest X-ray Image data set and produced state-of-the-art results. We demonstrate how to apply the proposed approach for more general big data problems by improving the explain ability and interpretability without sacrificing the prediction performance. The contribution of this paper opens a novel angle that moves the community closer to the future pipelines of XAI problems.