

1 Abstract

Machine learning has been a major player in the role of audio processing and classification for decades. In this time, a great deal of work has been devoted to studying the performance of various model architectures and producing sets of features that can represent a waveform in a compact, efficient, and non-redundant way [Citation?](#). In this work, we show that a spectrogram matrix and a feature-vector with predictors derived from both time-series and frequency-series representations of audio can be used to map that waveform to a potential source. This combination of features warrants a model design that combines a convolutional neural network (CNN) and a multilayer perceptron (MLP) to process the spectrogram matrix and the feature-vector respectively. We detail the significance and behavior of branch of the network and explore how this hybridization architecture along with the chosen features produces improved classification performance while retaining computational practicality.