1 Abstract

Machine learning algorithms have been a major performer 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 a set of features that can represent a waveform in a compact, efficient, and non-redundant way Citation?. In this paper, we explore a hybrid convolutional neural network (CNN) and multilayer perceptron (MLP) model that uses time-series and frequency-series features from a waveform to map it to a potential source. We detail the features used in classification as well as their physical significance as well as how the choice of feature influences the neural network architecture. In doing so, we show how the appropriate combination of architecture and features can produce reasonable classification performance when mapping sound files to a musical instrument source.