ABSTRACT

GOMES JUNIOR, Sergio Pinto. *Speech Emotion Recognition using Transfer Learning.* 99 f. Dissertação (Mestrado em Engenharia Eletrônica) - Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2019.

Speech has become an increasingly important mean of interaction between humans and computers. Aiming at making this interaction even more natural, researchers have proposed different systems of speech emotion recognition. In this area, in recent years, deep neural networks have been the focus of research. Given this, in this work we investigated the effect of techniques of transfer learning and data augmentation on the accuracy obtained by a residual convolutional neural network, comparing it to other classification strategies like ResNet without pre-training, the Gaussian Mixture Model and the Probabilistic Neural Network. In addition, samples of the Happiness, Neutral, Anger and Sadness classes contained in two emotion datasets (IEMOCAP and EmoDb) were used for training and testing of the proposed systems. In the experiments with the GMM, a accuracy of 85.77 % was achieved for the EmoDb dataset, and 66.83 % for the IEMOCAP. On the other hand, the probabilistic network developed in this work was able to correctly classify 79.64 % of the test samples from the EmoDb dataset. In the experiments with ResNet, the spectrograms of the speech signals were generated to be used instead of images. In these experiments it was observed that the techniques of data augmentation and transfer learning greatly contribute to the result of the emotion recognition. Using those techniques, the convolutional network correctly classified 81.26 % of the test samples.

Keywords: Emotion Recognition; Speech; Convolutional Neural Networks; Transfer Learning.