

Performance of machine learning classifiers in distress keywords recognition for audio surveillance applications

ABSTRACT

The ability to recognize distress speech is the essence of an intelligent audio surveillance system. With this ability, the surveillance system can be configured to detect specific distress keywords and launch appropriate actions to prevent unwanted incidents from progressing. This paper aims to find potential distress keywords that the audio surveillance system could recognize. The idea is to use a machine learning classifier as the recognition engine. Five distress keywords: 'Help', 'No', 'Oi', 'Please', and 'Tolong' were selected to be analyzed. A total of 515 audio signals comprising these five distress keywords were collected and used in the training and testing of 27 classifier models, derived from the Decision Tree, Naïve Bias, Support Vector Machine, K-Nearest Neighbour, Ensemble, and Artificial Neural Network. The features extracted from each audio signal are the Mel-frequency Cepstral Coefficients, while the Principal Component Analysis was applied for feature reduction. The results show that the keyword 'Please' is the most recognized, followed by 'Help', 'Oi', 'No' and 'Tolong', respectively. This observation was achieved using the Ensemble Bagged Trees classifier, which can recognize 'Please' with 99% accuracy in training and 100% accuracy in testing.