What is the aim?

The aim of the work is automatic speech recognition for speech test of Persian preschool children, the demand on what grew up rapidly since COVID-19.

Which model has been used?

1) Wav2Vec 2.0 model with modifying for various frequency domains.

2) VAD + Classifier

Which dataset has been used?

1) Data collected by asking some adults from social media and some students from an elementary school to participate in the experiment.

2) LibriSpeech, which s a collection of approximately 1,000 hours of audiobooks that are a part of the LibriVox project.

3) Common Voice is an audio dataset that consists of a unique MP3 and corresponding text file.

What is the result?

Following approach reaches a Word Error Rate(WER) of 6.45 on the Persian section of the CommonVoice dataset.

Pre-train Objective	Dataset	WER
RFP	Common Voice Persian	6.45
Masking	Common Voice Persian	8.45
RFP	LibriSpeech test–clean	1.35
Masking	LibriSpeech test–clean	1.79
RFP	Common Voice + RAN's Samples	4.56
Masking	Common Voice + RAN's Samples	5.15
RFP	Common Voice + Meaningless Words's Samples	4.12
Masking	Common Voice + Meaningless Words's Samples	7.15