Proceedings of the Sixth International Conference on Trends in Electronics and Informatics ICOEI 2022
DVD Part Number: CFP22132-DVD; ISBN: 978-1-6654-8327-8

Speech Emotion Recognition using Convolutional Neural Networks and Long Short-Term Memory

Ved Kokane
Dept. of Computer Engineering
Vidyavardhini's College of Engineering and Technology
Vasai, India,
ved.181403101@vcet.edu.in

Prasad Nijai
Dept. of Computer Engineering
Vidyavardhini's College of Engineering and Technology
Vasai, India,
prasad.173109151@vcet.edu.in

Vikas Jamge
Dept. of Computer Engineering
Vidyavardhini's College of Engineering and Technology
Vasai, India,
vikas.181383107@vcet.edu.in

Dr. Tatwadarshi P. Nagarhalli
Dept. of Computer Engineering
Vidyavardhini's College of Engineering and Technology
Vasai, India,
tatwadarshi.nagarhalli@vcet.edu.in

***Abstract*— In daily interpersonal human relationships, emotion plays a crucial role. This is necessary for both sensible and intelligent decisions. By expressing our sentiments and providing feedback to others, it helps us match and comprehend the feelings of others. The task of Speech Emotion Recognition (SER) is to recognise the emotion from speech regardless of the semantic content. Emotions, on the other hand, are subjective, and even humans find it difficult to record them in natural spoken communication, regardless of their meaning. The ability to conduct it automatically is a demanding endeavour that is continuously being researched. Emotion has a significant impact on human social interaction, according to research. Emotional displays reveal a great deal about a person's mental condition.**

Keywords—Machine Learning, Recurrent Neural Networks, Emotions, Long Short-Term Memory.

I. Introduction

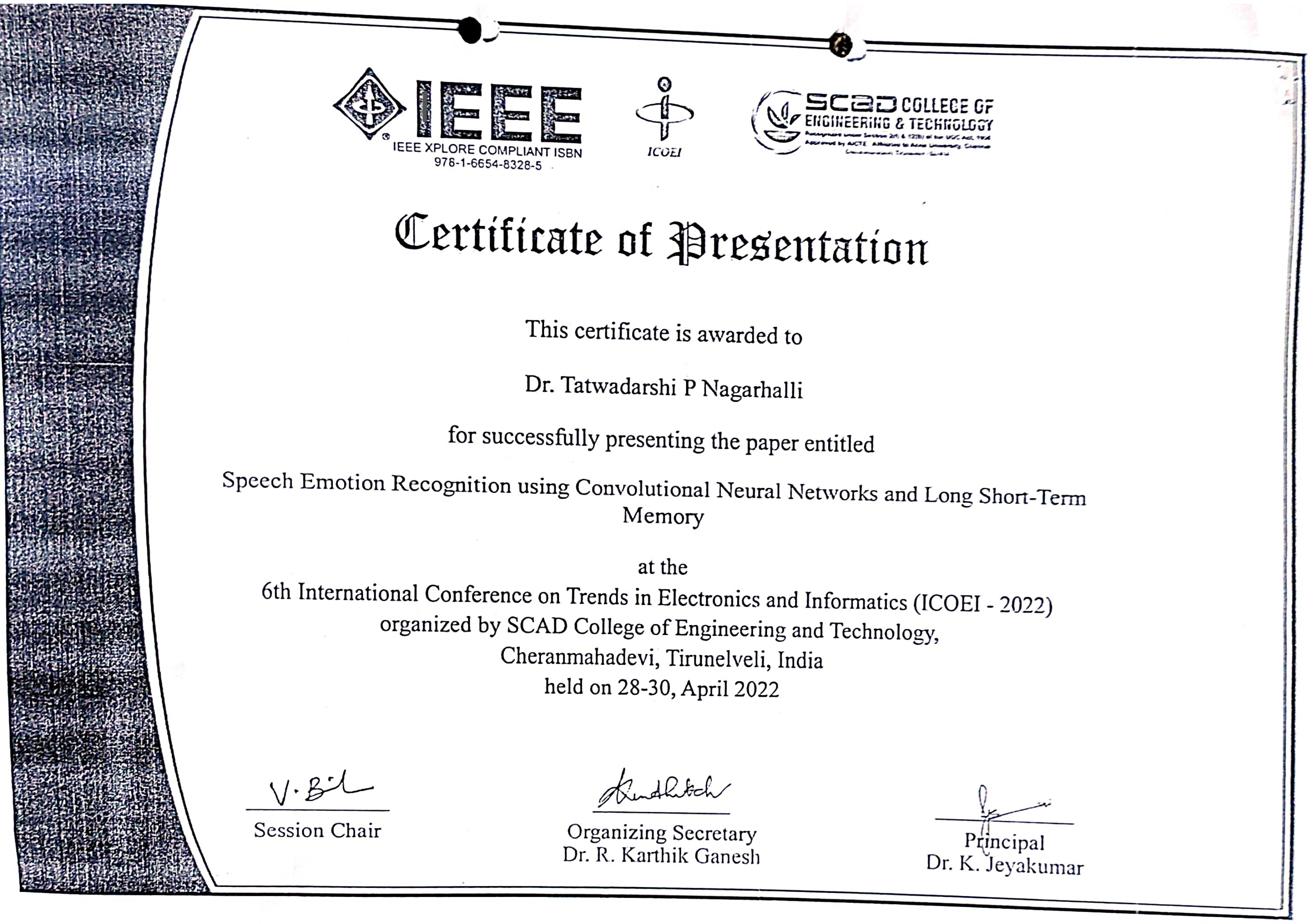
Perceiving emotions is consequently and subliminally performed by people. It is a crucial cycle for human-to human correspondence, and consequently, to accomplish better human to machine association, feelings should be thought of. Feeling acknowledgment from a human discourse is an alluring field of discourse signal preparation. It is attracting more attention the applications where feeling acknowledgment facilitates the speaker's recognizable proof and mental status, for example, in criminal examination, clever help, identifying frustration, frustration, shock/entertainment, medical care furthermore, medication and a superior Human Computer Interface. Separating the enthusiastic condition of a speaker from his/her discourse is called speech emotion recognition.

The way of speaking is the quickest and most distinctive method of communication between persons. This reality has led experts to regard speaking as a quick and effective means of communication in a crowd of people acting like robots. On the other side, this necessitates the machine's ability to recognise human voices. Since the late 1950s, there has been extensive research into the recognition of discourse modes, which indicates progress in converting human speech into a collection of words. Regardless, despite tremendous advancements in voice recognition, we are still a long way from having a trademark relationship between man and machine, because the computer does not understand the speaker's excited state.

Recognition of the emotional condition is a huge element of human-machine interface research. Ascribes used in acknowledgment of feeling were ensuing from changes in facial copies just as discourse signs. Feeling remains as a physiological reaction that follows in circumstances like misery, dread or bliss. Essentially, age of discourse is a physiological cycle where the variety of conditions of feeling will be imitated in discourse just as face. This change in speech will likewise be doled out of personality, mental status and actual wellbeing of an individual.

Speech Emotion Recognition is a field that is constantly evolving. The use of machine learning and deep learning in the field of sound investigation is rapidly growing. A couple of versions include pre-programmed discourse recognition, computerised signal processing, and sound order, labelling, and age. Menial helpers like Alexa, Siri, and Google Home have spent a lot of time developing algorithms that can detect fake information from sound.

Feature extraction employs a variety of audio features. The most extreme amplitude values among all examples in each frame make up the signal's Amplitude Envelope. This element conveys a sense of volume. It is, nevertheless, vulnerable to abnormalities. This component has been widely used in the identification and classification of music genres. All samples in a frame are used to calculate Root Mean Square Energy.It is used to indicate volume, as the more the energy, the louder the sound. It is, however, less vulnerable to outliers than the Amplitude Envelope. Audio segmentation and music genre classification have both benefited from this capability. The zero-crossing rate is the number of times a waveform crosses the horizontal time axis. This feature has been frequently utilised.



IEEE XPLORE COMPLIANT ISBN
978-1-6654-8328-5

ICOEI

SCAD COLLEGE OF ENGINEERING & TECHNOLOGY
Recognised under Section 2(f) & 12(B) of the UGC Act, 1956
Approved by AICTE. Affiliated to Anna University, Chennai
[illegible]

Certificate of Presentation

This certificate is awarded to

Dr. Tatwadarshi P Nagarhalli

for successfully presenting the paper entitled

Speech Emotion Recognition using Convolutional Neural Networks and Long Short-Term Memory

at the
6th International Conference on Trends in Electronics and Informatics (ICOEI - 2022)
organized by SCAD College of Engineering and Technology,
Cheranmahadevi, Tirunelveli, India
held on 28-30, April 2022

Session Chair

Organizing Secretary
Dr. R. Karthik Ganesh

Principal
Dr. K. Jeyakumar

HEAD
Dept of Computer Engg.,
Vidyavardhini's College of
Engineering and Technology,
Vasai Road 401 202

VIDYAVARDHINI'S COLLEGE OF ENGINEERING & TECHNOLOGY *
COMP
Engg.