

# Emotion recognition from audio-visual big data using Deep Learning Techniques

Vinay R. Kurmi

## Introduction

Human emotion play a vital role in narrating a story or a life event. Classification of the emotional category within an audio-visual stream using robust deep learning techniques to improve customer experience and consumer feedback system.

## Objectives

- Identifying suitable methodologies for emotion classification in real-time audio-visual stream ?
- How demographics information collaboratively with audio-visual data improve the speech emotion recognition system ?
- Technical possibilities to build a online system for emotion classification ?

- Eric Chu and Deb Roy constructed method of classifying the families of emotional arcs, which displays predictive power for emotional arcs of short Vimeo videos and propose to implement their machine learning techniques to propagate through social media like Facebook, Tiktok, Instagram Twitter and Reddit

## Proposed Methodology

- Data Collection: Publicly available Youtube, Spotify, Vimeo, Soundcloud and other opensource databases would act as the training data for this process.
- Segmentation of the available audio and video streams
- Artificial Neural Network, Deep Learning, Convolutional Neural Network, Deep Convolutional Neural Network and LSTM Networks have proven to be some of the most effective and recent speech emotion recognition techniques
- Deep convolutional neural networks is one of the prominent machine learning techniques to process audio-visual data. Past research proposes utilization of Acoustic Emotion Recognition using LSTM.

## Challenges

- Emotions with a audio stream is frequently dependent on the visual graphics of the scene.
- Background noise reduction using machine learning algorithms can prove to be a challenging task.
- Phycologists have been examining the association of emotion along with facial expression since mid 90's, yet they still remain to fill all gaps in such an experiment. This study aims to aid their goal of emotion recognition.

## Results & Benefits

- Improvement in accuracy of real-time emotion recognition from audio-visual data.
- Enhancing customer experiences with real-time analysis.
- Social Media Exposure through emotion classification within events.

## Potential Artefact

- Improve robustness of the Deep Learning Techniques.
- Additional challenge that needs to be resolved is processing of overlaps between speakers and crowd emotional manifestation which helps us with acoustic cues classifying group and crowd vocal manifestation in abnormal situations.
- Enhance the accuracy of the recently used deep learning techniques for emotion recognition.

## Use Cases

- Digital Advertisement
- Online Gaming
- Customer Feedback Assessment
- Healthcare
- Mental Health Analysis
- Effective Learning System
- Artificial Intelligence
- Computer-Human Interaction
- Entertainment Industry
- Virtual Assistant
- Match Making
- Deep Fake Detection

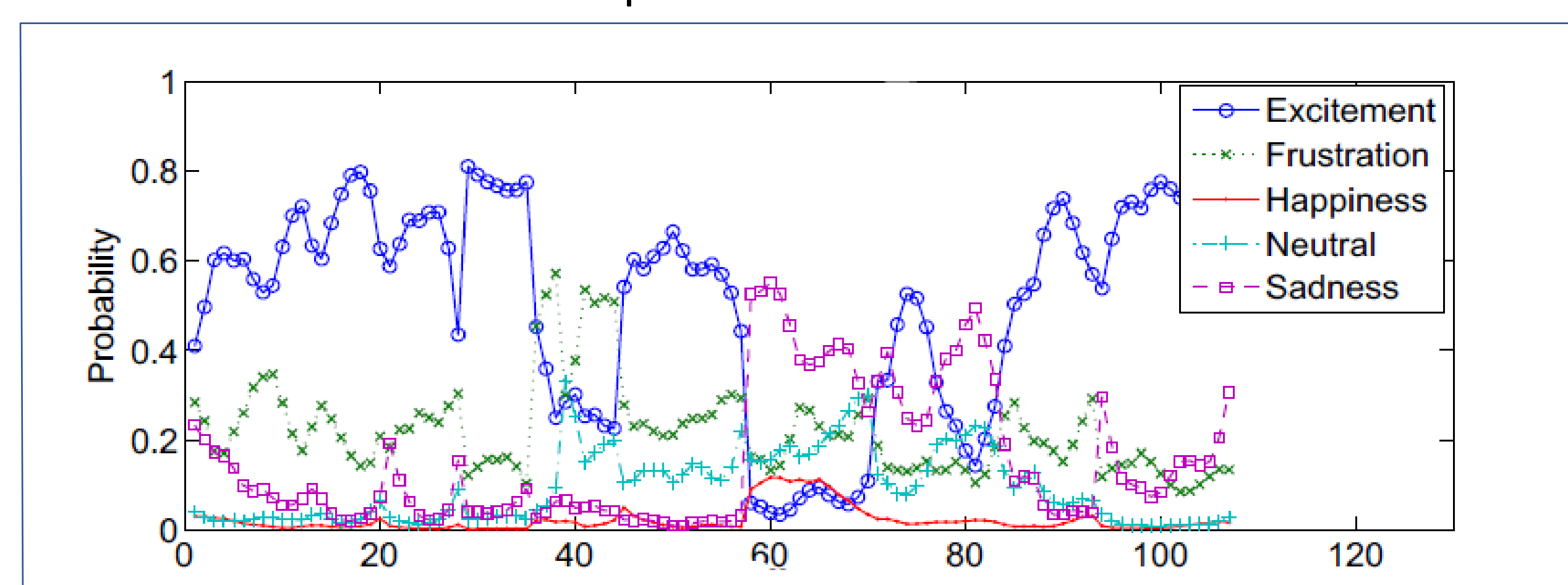
## Overview

- M. Shamim Hossain and Ghulam Muhammad used 2D CNN for audio and 3D CNN for video signals for evaluation of Big data using SVM classifier and the proposed system had an 99.9% accuracy in emotion recognition.
- Jaejin Cho combined acoustic features and transcripts in his research to improve emotion recognition, this system improved the recognition by 18% which proves that acoustic and transcripts compliment each other for emotion recognition.

## Knowledge Area Points

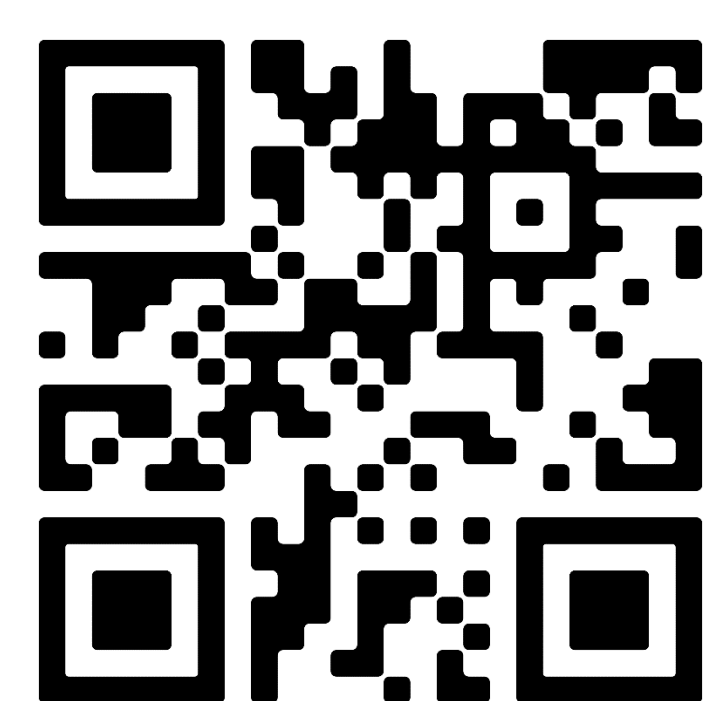
- Emotional Arcs
- Sentiment Analysis
- Deep Learning Techniques
- Speech Emotion Recognition
- Deep Neural Networks
- Signal Processing
- Probabilistic Modelling
- Regression
- Machine Learning
- Convolutional Neural Networks
- Classification
- Image Processing

Figure 1. DNN Outputs for each segment of the audio-visual where each line depicts an emotional state.



## Contact

Vinay R. Kurmi  
Dublin Business School  
10576078@mydbs.ie  
+353 0894207666



## References

1. Babak Joze Abbaschian, D. S.-S. A. E., 2021. Deep Learning Techniques for Speech Emotion Recognition from Databases to Models. sensors, MDPI, Volume 1, p. 27.
2. Eric Chu, D. R., 2017. Audio-Visual Sentiment Analysis for Learning Emotional Arcs in Movies. 2017 IEEE International Conference on Data Mining, Volume 1, p. 6.
3. Jaejin Cho, R. P. K. J. V. Y. C. N. D., 2019. Deep neural networks for emotion recognition combining audio and transcripts. Center for Language Speech Processing, Johns Hopkins University, Baltimore, MD, USA, Volume 1, p. 5.
4. Kun Han, D. Y. I. T., 2014. Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine. INTERSPEECH 2014, Volume 1, p. 5.
5. Lakshmin Kaushik, A. S. J. H. L. H., 2013. SENTIMENT EXTRACTION FROM NATURAL AUDIO STREAMS. ICASSP2013, Volume 1, p. 5.
6. M. Shamim Hossain, G. M., 2018. Emotion Recognition Using Deep Learning Approach from Audio-Visual Emotional Big Data. Information Fusion (2018), Volume 1, p. 24.
7. Martin Wöllmer, F. W. T. K. a. B. S. T., n.d. YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context. IEEE INTELLIGENT SYSTEMS, Volume 1, p. 8.