International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 02 | February - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

Multi-modal Emotion Recognition Using Neural Networks: Deploy on Cloud

Rashmi K[1], Shreya B S[2], Adithi S Athreya[3], Reethu M[4], Swathi K[5]

[1]CSE(AI & ML), Vidyavardhaka College of Engineering
[2]CSE(AI & ML), Vidyavardhaka College of Engineering
[3]CSE(AI & ML), Vidyavardhaka College of Engineering
[4]CSE(AI & ML), Vidyavardhaka College of Engineering
[5]CSE(AI & ML), Vidyavardhaka College of Engineering

Abstract:

This paper explores recent developments in emotion recognition, focusing on methodologies employing Multilayer Perceptron (MLP) classifiers. Emotion recognition is crucial for human-machine interaction across various domains, including healthcare, surveillance systems, and intelligent learning environments. There are many deep learning models that have been employed for emotion detection, this emphasizes the effectiveness of MLP classifiers. The project at hand centers on predicting emotions such as happiness, calmness, fear, and disgust solely utilizing MLP classifiers. The system is trained on audio data, with recordings matched in frequency, time, and speed to the raw data used for training. Additionally, Convolutional Neural Network (CNN) models are used here to analyze facial expressions in real-time video streams by providing neat and complete approach to emotion recognition. Upon deployment on the AWS cloud platform, the system provides real-time emotion predictions in textual format via a web interface created using Flask. Furthermore, the system offers the capability to analyze emotions from live video streams, presenting continuous feedback on emotions through a face-outlined box updated every three seconds. This functionality serves to confirm individuals' emotions and provide self-awareness regarding their emotional states. This project focuses on giving efficiency by using MLP classifiers in conjunction with CNN models for emotion recognition tasks, thereby contributing to the advancement of human-machine interaction.

Key Words: MLP Classifier, CNN, AWS Cloud, Flask, emotion, audio, video

1. INTRODUCTION:

Emotion recognition is a vital aspect of human-machine interaction, with applications spanning diverse domains such as healthcare, surveillance systems, and intelligent learning environments. There are several deep learning models that used for emotion detection purpose, here it focuses on Multilayer Perceptron (MLP) classifiers. MLP classifiers have shown promise in accurately predicting emotions from various modalities, including Verbal expressions, facial reactions and sound signal inputs. However, it's important to make sure that models, such as Convolutional Neural Networks (CNNs) and Deep Belief Networks (DBNs), have also been employed for Emotion recognition tasks. The project discussed here utilizes MLP classifiers exclusively for predicting emotions from audio data and is deployed on the AWS cloud platform. Nonetheless, the findings presented herein may offer insights into the potential applications and effectiveness of MLP classifiers when compared to other models. Additionally, the project integrates CNN models for real-time analysis of facial expressions in video streams. This comprehensive approach aims to provide nuanced insights into emotional states, thereby advancing the field of human-machine interaction. Through this exploration, we aim to contribute to the ongoing discourse surrounding the choice of models in emotion recognition system.

2. METHODOLOGY

The developed system is designed to accept input from both audio and video sources, initiating several processes driven by augmentation techniques and innovative deployment strategies. It delves into the domain of computerized emotion recognition from both audio and video data, exploring diverse methodologies and difficulties that are naturally part of the process. While emphasizing the performance of Multilayer Perceptron (MLP) classifiers for audio-based emotion recognition, the methodology also incorporates techniques tailored for video data processing through pre-trained CNN models. These techniques enable the system to analyze facial expressions in conjunction with audio cues, thereby enhancing its capacity to comprehend human emotions across multiple modalities. The proposed approach spans various stages, including data preparation, model training, evaluation, and integration into web interfaces, with the overarching objective of advancing computer understanding of human emotions for real-world applications.

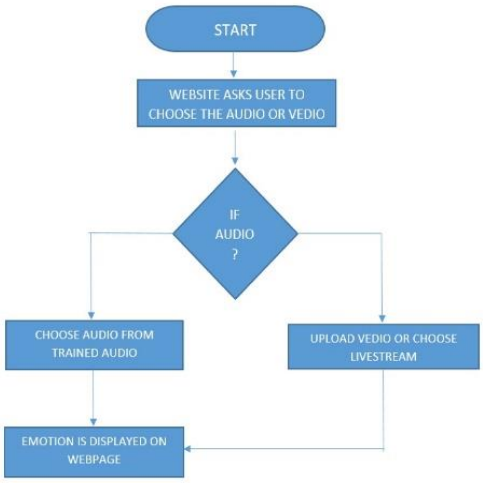

Fig 1. Implementing a mechanism that combines the result of both audio and video emotion recognition.

2.1 AUGMENTATION PROCESS –AUDIO:

To ensure compatibility and consistency in input data, an augmentation process is implemented. This process involves matching the characteristics of raw audio data with clean reference audio, equalizing amplitude, and adjusting duration as necessary. The augmentation process enhances the quality and uniformity of input data, improving the accuracy and reliability of emotion recognition.

2.2 VIDEO PROCESSING:

The video live streaming emotion detection system processes real-time video streams using OpenCV to capture frames. It captures video frames from the webcam, detects faces in those frames, and predicts the emotions associated with those detected faces using the loaded model.

It applies a face detection algorithm to locate faces, then resizes and normalizes them before feeding them into a pre-trained convolutional neural network (CNN) model for emotion classification. Predicted emotions are overlaid on the video frames as text. For static images, a similar process is followed, but the system operates on individual images instead of continuous video streams. This approach enables real-time emotion detection in videos and images, facilitating applications such as human-computer interaction and sentiment analysis.

2.3 WEBSITE FUNCTIONALITY:

The project is deployed on a web interface, providing users with the capability to input audio and video data for emotion recognition. Upon selection or recording of audio and live stream video, the system triggers the augmentation process and executes emotion recognition algorithms using MLP classifiers and CNN. The results are displayed in real-time using the interface on the internet, enabling users to interpret and analyze emotional states effectively.

2.4. CLOUD DEPLOYMENT:

To ensure scalability, accessibility, and real-time analysis, the project is deployed on the Amazon Web Services (AWS) cloud platform. Leveraging the computational resources and infrastructure provided by AWS, the system achieves efficient execution and analysation of audio and video data, facilitating seamless integration and deployment in diverse applications and environments.

3. CONCLUSION:

As referred to the limitations in existing literature, the project presents a compelling solution to address key challenges in emotion detection systems. Despite the shortcomings observed in traditional approaches such as recurrent neural networks (RNNs) and the limited evaluation criteria encompassing only four basic emotions, the project stands out by leveraging Multilayer Perceptron (MLP) classifiers to achieve robust emotion recognition capabilities.

One prominent issue highlighted in this review is the lack of coverage of long context information in RNNs, leading to difficulties in capturing temporal dependencies effectively. The project sidesteps this challenge by utilizing MLP classifiers, which are capable of processing input data simultaneously, therefore mitigating the gradient vanishing problem encountered in RNNs.

Moreover, while previous implementations have struggled with real-time deployment on resource-constrained devices like Raspberry Pi, the project demonstrates feasibility by successfully deploying the system on such devices through cloud infrastructure. Despite initial limitations in recording and validating only four expressions, the project showcases scalability and adaptability, paving the way for future implementations to further enhance system efficiency.

Furthermore, this project overcomes challenges related to memory usage and dataset limitations encountered in previous studies. By optimizing resource utilization and processing both audio and video files efficiently, reliable performance is achieved without compromising on system stability or scalability.

This project offers a significant advancement in emotion recognition systems by providing a robust, scalable, and efficient solution that addresses key limitations identified in existing literature. By leveraging MLP classifiers, CNN and innovative deployment strategies, it demonstrates the potential to overcome traditional challenges and pave the way for more effective human-machine interaction in diverse applications as a multi-model system.

REFERENCES:

1. Mandeep Singh SCPD, Stanford University Stanford, CA, et al., Emotion Recognition in Audio and Video Using Deep Neural Networks, 15 june 2020
2. Shiqing Zhang, Shiliang Zhang, Member, IEEE, et al., Learning Affective Features With a Hybrid Deep Model for Audio–Visual Emotion Recognition , volume 28 , october 2018
3. Marwan Dhuheir, Abdullatif Albaseer , et al., Emotion Recognition for Healthcare Surveillance Systems Using Neural Networks: A Survey, 13 july 2021
4. Humaid Saif Alshamsi, et al. Automated Facial Expression and Speech Emotion Recognition App Development on Smart Phones using Cloud Computing, November 2018
5. Zeenat Tariq, University of Missouri, et al. Speech Emotion Detection using IoT based Deep Learning for Health Care, December 2019
6. Navjot Rathour et al., IoMT Based Facial Emotion Recognition System using Deep Convolution Neural Network.,28 May 2021
7. Md Zia Uddin et al., Emotion Recognition using Speech and Neural structured learning to facilitate edge intelligence., 24 June 2020
8. Ganesh K Venayagamorthy Voice Recognition using neural networks,3 Feb 2016
9. Anushka Sandesara et al., A Cooperative Study On Speech Emotion Recognition., Volume 3, November 2020
10. STMIK Dharma Wacana., Human Voice Recognition using Neural Networks., 20 April 2021
11. Hadhami Aouani et al., Speech Emotion Recognition using deep learning., 2020
12. Sukpreeth Kaur et al., Emotion Recognition – A Review., volume 16., November 2 2021
13. Xiofeng Lu et al., Deep learning based emotion recognition and visualization of figural representation., volume 12., January 2022
14. Nagaraja N Poojary et al., Speech emotion Recognition using MLP classifier., volume 7., 15 july 2021