# INTRODUCTION

**Analyzing Emotions in Audio: Using Machine Learning to Classify Human Emotional States**

In the realm of sound-based recognition projects, determining human emotional states presents a unique challenge. Adopting a discrete emotional approach, we delve into emotions like anger, boredom, disgust, surprise, fear, joy, happiness, neutral, and sadness.

Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) offer efficient solutions for image and video processing, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks shine in speech-based classification, such as Natural Language Processing (NLP) and Speech Emotion Recognition (SER).

Speech Emotion Recognition involves extracting emotion features from speech signals, leveraging parameters like pitch, volume, and energy. However, each model has its limitations. While CNNs excel in learning features from high-dimensional input data, they demand significant storage. Conversely, LSTM-based RNNs handle variable input data and sequential text data effectively.

**PROBLEM STATEMENT**

This method required enormous engineering features and any variation in the features would need re-modeling the overall architecture of the technique. Nevertheless recent development in deep learning applications and methods for Speech Emotion Recognition can be varied .There are numerous literature and studies on the application of these algorithms to understand emotions and state of mind from human speech.

Additionally to deep learning, neural networks and application of improvements of long short term memory (LSTM) networks, generative adversarial models, and lots more a wave in research on speech emotion recognition and its application now emerges. It is essential to understand its application and its role in emotion. For this reason, the objective of the project is to understand deep learning techniques for speech emotion recognition from databases to models.

It is a challenge to make emotion available in different languages. There are limitations with different types and versions of the software’s such as dataset input is only textual data, image, pattern, video and audio inputs are invalid. Performance and results of the emotion sensing system depends on accuracy of the sensors such as cameras, thermal image sensors, facial recognition algorithm used and so on.

Highly accurate system will be expensive due to use of costly components. After applying the deep learning feature extraction methods further, it is very difficult to obtain high accuracy in the model because of the similarities between the different emotions like happy and surprising. As these emotions have the same kind of frequency and tone. The length of the voice is also a problem because that the human emotions do not remain the same throughout the sentence it keeps on changing so the system has to identify the parts of the data to understand the full emotion of the voice.