Emotion Detection From Urdu Speech

Shayaan Qazi Ikhlas Ahmed Sameer Kamani Hunain Abbas Dr. Abdul Samad Sandesh Kumar

Habib University
shaping futures

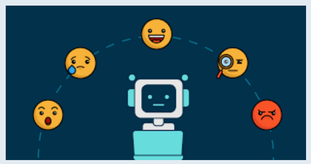
Introduction & Research Question

Detecting emotions from speech often relies on tonal and acoustic cues. While English-based models exist, Urdu's unique phonetics and lack of labeled data pose challenges. This project has worked on developing accurate emotion detection models for Urdu speech using advanced deep learning and transformer techniques.

Research Question: Given an Urdu audio clip, how accurately can we classify the speaker's emotion?

Motivation

Empathetic AI-chatbots



Ethical Dilemma of AI Chatbots

Urdu recognizing AI Lagging Behind for 250M+ Speakers

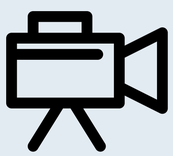
Pakistan Population (LIVE)

Data Set

Total actors
24

~10,000 training

Total Recordings
14,000+

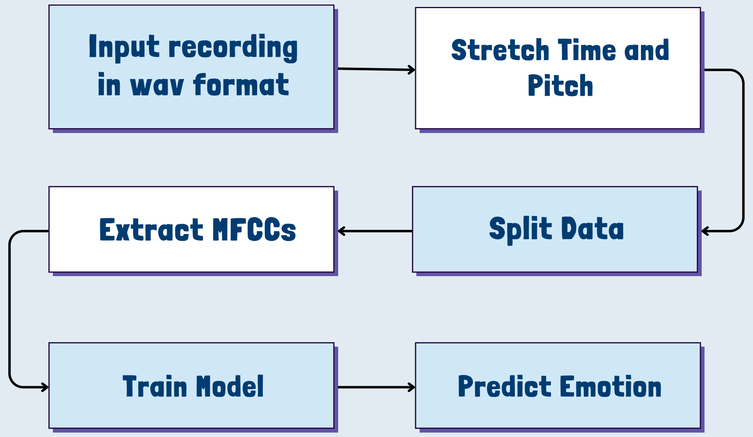


~1,000 Validation

Source
SEMOUR+

~3,000 testing

Methodology



Results

Model	Result
SVM	62%
CNN	85.09%
Resnet50	86.82%
Hubert	90%
KNN	91.16%
Wav2vec2.0	94.5%

Figure 1. Accuracy of all the models we trained, from worst to best.

Wav2Vec 2.0 Confusion Matrix

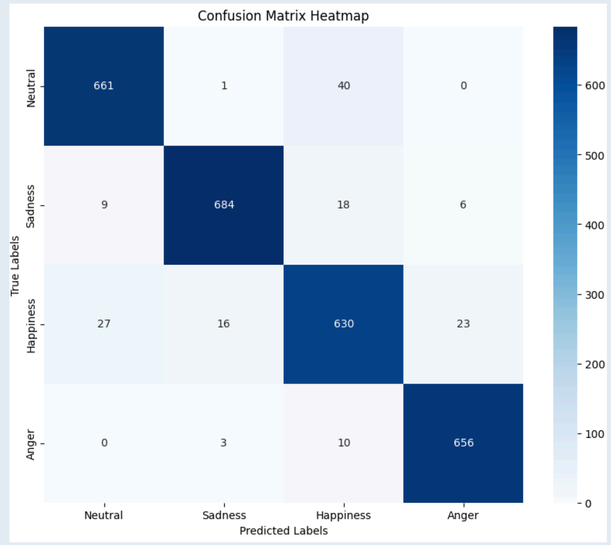


Figure 2. Confusion Matrix depicting the model's correct and incorrect predictions for each emotion.

Related Work

Their best
KNN - 82.5%

Our best
WAV2VEC 2.0 - 94.5%

Table 8 Comparison with related work.

Papers	Languages	Training technique	Features extraction techniques	Emotions	Classifier used	Accuracy
Tripathi & Beigi (2018)	English and German	Speaker dependent	RNN	Anger, happiness, neutral and sadness	RNN with three layers	71.04%
Kamińska, Sapiński & Anbarjafari (2017)	Polish	Speaker dependent independent	MFCC, BFCC, RASTA, energy, formants, LPC and HFCC	Sadness, happiness, anger, neutral, joy, fear and surprise	SVM and k-NN	81%
Rajisha, Sunija & Riyas (2016)	Malayalam	Speaker dependent	MFCC, STE and pitch	Neutral, anger, happiness and sad	ANN and SVM	78%
Ali et al. (2013)	Urdu	Speaker dependent	Duration, intensity, pitch and formants	Anger, sadness, happiness and comfort	Neive Bayes	76%
Abbas, Zehra & Arif (2013)	Urdu	Speaker dependent	Intensity, pitch and formants	Anger, sadness, happiness and comfort	SMO, MLP, J48 and Neive Bayes	75%
Latif et al. (2018)	Urdu	Speaker independent	LLDs low level descriptor	Happiness, sadness, anger and neutral	SVM, logistic regression and RF	83%
Sinith et al. (2015)	English Malayalam and	Speaker dependent	MFCC, pitch and energy	Anger, neutral sadness and happiness	SVM	70%
Our work	Urdu (with disgust emotion)	Speaker dependent	MFCC, LPC, energy, pitch, zero crossing, spectral flux spectral centroid, spectral roll off	Anger, disgust, happiness, sadness and neutral	k-Nearest Neighbours	73%
Our work	Urdu (without disgust emotion)	Speaker dependent	MFCC, LPC, energy, pitch, zero crossing, spectral flux spectral centroid, spectral roll off	Anger, happiness, sadness and neutral	k-Nearest Neighbors	82 .5%

(Asghar, Sohaib, Iftikhar, Shafi, & Fatima, 2022)

Conclusion and Plans for the Future

- We managed to train a model that generalises well and is highly accurate.
- However, we can improve:
  - Newer feature extraction methods,
  - Better data augmentation,
  - Work on more emotions.