DON BOSCO INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

UNDER DBIT - SIC

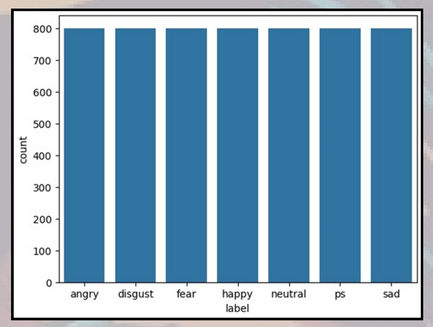
Together for Tomorrow! Enabling People
Education for Future Generations

PROBLEM STATEMENT

Develop a speech emotion recognition system utilizing LSTM and CNN algorithms to accurately classify emotions from audio data, enhancing human-computer interaction, sentiment analysis, and mental health applications.

DATA SOURCE

The dataset used is considered from KAGGLE, the TESS Toronto emotional speech set data which has 14 directories.



ANALYSIS

- Data Preparation
- Data Visualization and Exploration
- Data Augmentation
- Feature extraction
- Modelling

VARIABLES

- Fear
- Angry
- Sad
- Happy
- Neutral
- Disgust
- Calm
- Surprise

FEATURE SELECTION

- Zero Crossing Rate
- Chroma_stft
- MFCC
- RMS(root mean square) value
- MelSpectogram to train our model

PREDICTION MODEL

By Building an LSTM-based model for SER, it is possible to stack multiple LSTM layers to capture long-term dependencies in the audio data. Also Convolutional Neural Networks (CNN) can be used before LSTM layers for better feature extraction from spectrograms.

	precision	recall	f1-score	support
angry	0.96	0.97	0.97	1484
disgust	0.97	0.95	0.96	1558
fear	0.96	0.97	0.96	1505
happy	0.96	0.95	0.96	1619
neutral	0.97	0.98	0.97	1558
sad	0.96	0.97	0.96	1478
surprise	0.98	0.97	0.97	528
accuracy			0.96	9730
macro avg	0.96	0.96	0.96	9730
weighted avg	0.96	0.96	0.96	9730

MODEL COMPARISION

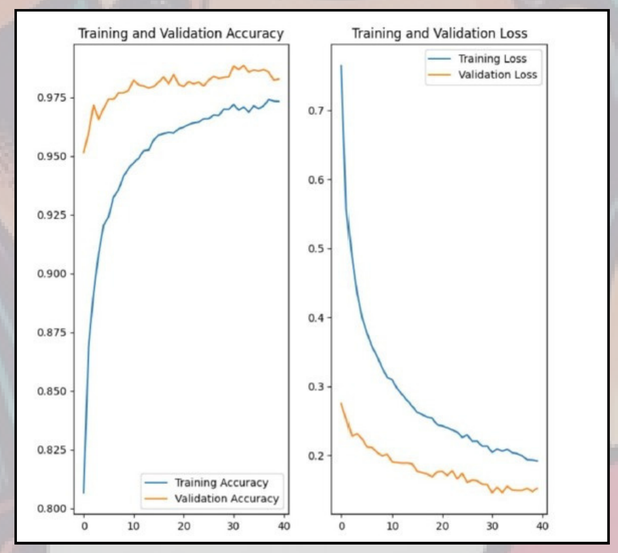
The Acurracy obtained from RNN Model

```
Epoch 10/10
9/9 [==============================] - 1s 63ms/step - loss: 0.4418 - accuracy: 0.8484 - val_loss: 0.2939 - val_accuracy: 0.9152
175/175 [==============================] - 1s 6ms/step - loss: 0.2718 - accuracy: 0.9307
Test Accuracy: 93.07%
```

The Acurracy obtained from LSTM Model

```
Epoch 9/100
9/9 [==============================] - 1s 59ms/step - loss: 0.0151 - accuracy: 0.9962 - val_loss: 0.0071 - val_accuracy: 0.9991
Epoch 10/100
9/9 [==============================] - 1s 60ms/step - loss: 0.0186 - accuracy: 0.9958 - val_loss: 0.0036 - val_accuracy: 0.9982
```

GRAPH



CONCLUSION

This project in speech emotion recognition using LSTM and CNN algorithms has been a significant contribution to human-computer interaction and affective computing. We aimed to create a robust system for identifying emotions from spoken language with broad applications in healthcare, customer service, and entertainment. The combination of LSTM and CNN algorithms proved powerful. LSTM captured temporal dependencies in audio data for nuanced emotion recognition, while CNN extracted features from spectrograms, focusing on pitch, tone, and intensity. Our experiments yielded highly accurate results, showcasing the effectiveness of our algorithms, dataset, and model design.

LOHITH KUMAR A - 1DB20CS066

RAKSHITHA S HIREMATH - 1DB20EC068