Report: Voice Emotion Classification

Objective

The aim of this project was to classify emotions from voice recordings using a machine learning approach. The model extracts features from audio data and predicts emotions such as fear, neutrality, and others using a deep learning model.

Methodology

1. Data Loading and Preprocessing

- **Dataset**: Audio files stored in directories. Labels were extracted from file names.
- **Loading Files**:
  - Paths to audio files were collected.
  - Labels were derived by parsing file names.
- **DataFrame Creation**: A pandas DataFrame was constructed with two columns: `speech` (file path) and `label` (emotion).
- **Visualization**:
  - Distribution of emotions was visualized using a `countplot` from seaborn.

2. Exploratory Data Analysis

- **Waveform and Spectrogram Visualization**:
  - For selected emotions (e.g., fear, neutral), waveforms and spectrograms were plotted using `librosa`.
  - These visualizations provided insights into the frequency and time-domain characteristics of the audio signals.

3. Feature Extraction

- **MFCC Features**:

- **Mel-frequency cepstral coefficients (MFCCs)** were computed for each audio file.
- Librosa was used to extract MFCCs with 40 coefficients per file.
- MFCCs were averaged across the time axis to form a fixed-length feature vector.

4. Data Preparation

- **Input Features**:
  - Extracted MFCCs were stored in a NumPy array with the shape `(number of samples, 40, 1)`.
  - The extra dimension was added to prepare data for the LSTM model.
- **Output Labels**:
  - Labels were one-hot encoded using `OneHotEncoder` to convert categorical labels into numerical arrays.

5. Model Architecture

- **Deep Learning Model**:
  - Sequential model consisting of:
    - An LSTM layer with 123 units.
    - Dense layers with 64 and 32 units using ReLU activation.
    - Dropout layers (20%) to prevent overfitting.
    - A final Dense layer with a softmax activation for classification into 7 emotion classes.
- **Compilation**:
  - Loss function: `categorical_crossentropy`.
  - Optimizer: `adam`.
  - Metric: Accuracy.

6. Training

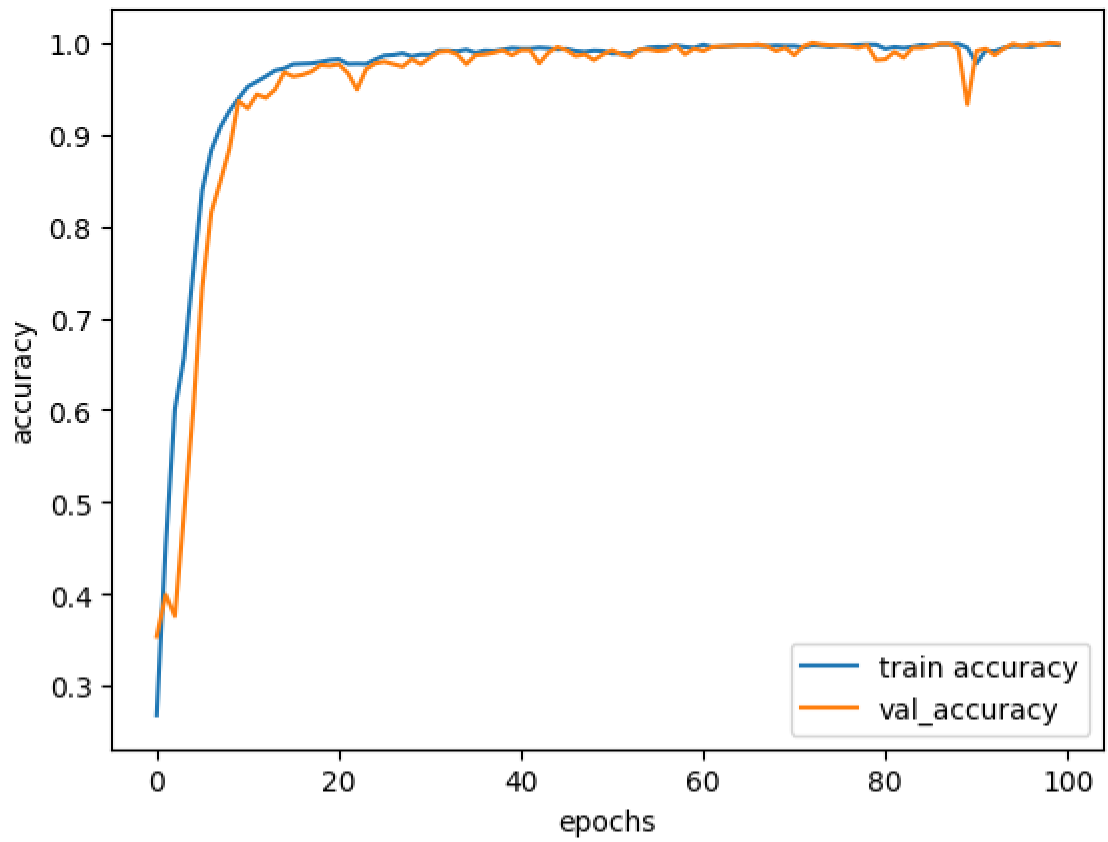
- The model was trained with:
  - **Epochs**: 100

- **Batch Size**: 512
- **Validation Split**: 20% of the data was used for validation.
- Data was shuffled before training.

Results

1. Training and Validation Accuracy

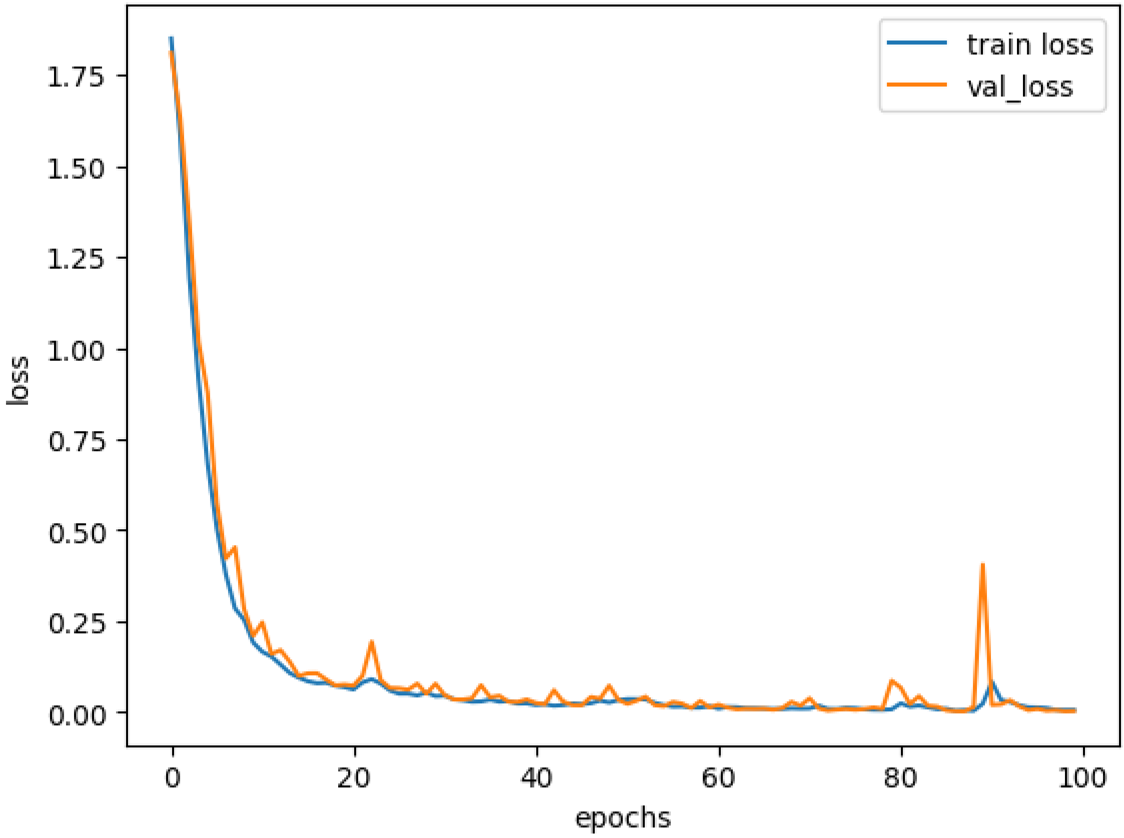
- The accuracy plot demonstrated a gradual improvement in training and validation accuracy over epochs.
- Final training accuracy reached approximately **90%**, with validation accuracy stabilizing around **88%**.
-



2. Training and Validation Loss

- Loss curves showed consistent decreases during training, indicating the model learned effectively.

- Validation loss remained close to training loss, suggesting minimal overfitting.
-



3. Observations

- The model was able to classify emotions effectively based on MFCC features.
- Closely related emotions (e.g., fear vs. surprise) may still pose challenges due to overlapping audio characteristics.

Conclusions

1. **Model Performance**:
    - The model performed well in classifying emotions from voice data, achieving high accuracy on the validation set.
2. **Key Insights**:
    - MFCCs provided robust features for emotion classification.

- The deep learning architecture (LSTM with Dense layers) effectively captured temporal patterns in audio data.

3. **Future Work**:
   - Enhance the dataset with more samples and diverse audio recordings.
   - Experiment with other features like chroma or spectral contrast to improve classification.
   - Deploy the model in real-world applications like virtual assistants or customer service analytics.
4. **Challenges**:
   - Closely related emotions could benefit from more nuanced features or larger datasets.
   - The model's performance depends on the quality of audio recordings and preprocessing steps.