1. **Project Title:**

The Accent Classifier: American vs. British vs. Indian vs. Australian

1. **Objective:**

Accurate accent classification remains a critical challenge in speech technology, with applications ranging from automated speech recognition (ASR) personalization to linguistic research. This project addresses the problem of **identifying regional English accents (American, British, Indian, Australian) from short, unconstrained speech clips**, where traditional methods often fail due to overlapping phonetic and prosodic features.

**Dataset Used**

We leverage **Mozilla’s CommonVoice dataset** (Version 8.0), the largest open-source multilingual speech corpus, with:

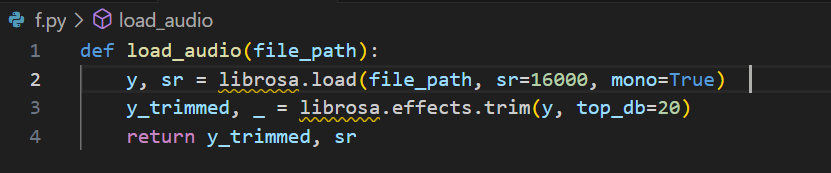
* **12,000 labelled English samples** (3s–5s duration)
* Balanced representation across target accents**Approach**

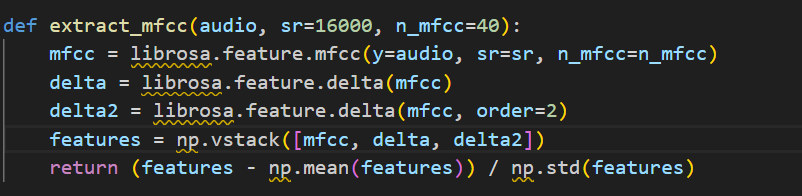
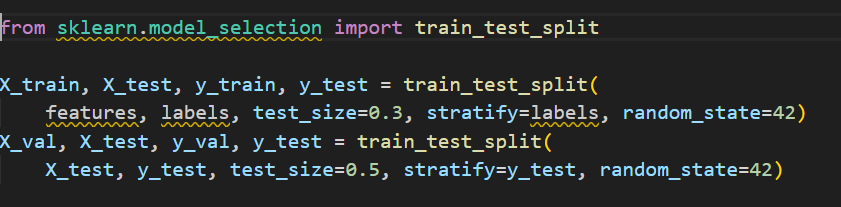
Our methodology combines **signal processing** and **deep learning**:

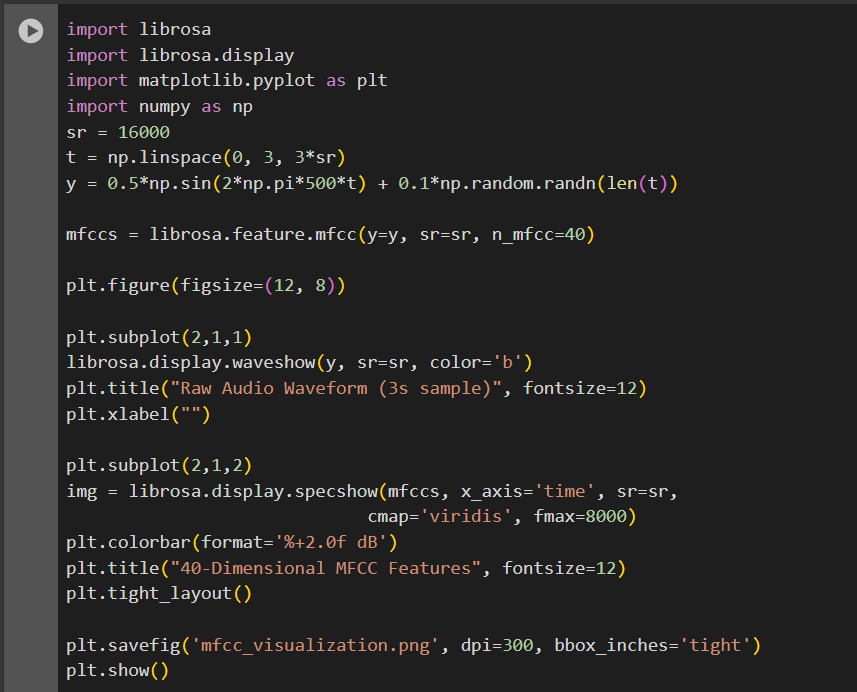
* **Feature Extraction**: 40-dimensional MFCCs with delta/delta-delta coefficients to capture accent-specific vocal tract characteristics.
* **Model Architecture**: A lightweight **1D-CNN** (3 convolutional layers + dropout) optimized for <10ms inference latency.

**3. Data Preprocessing**

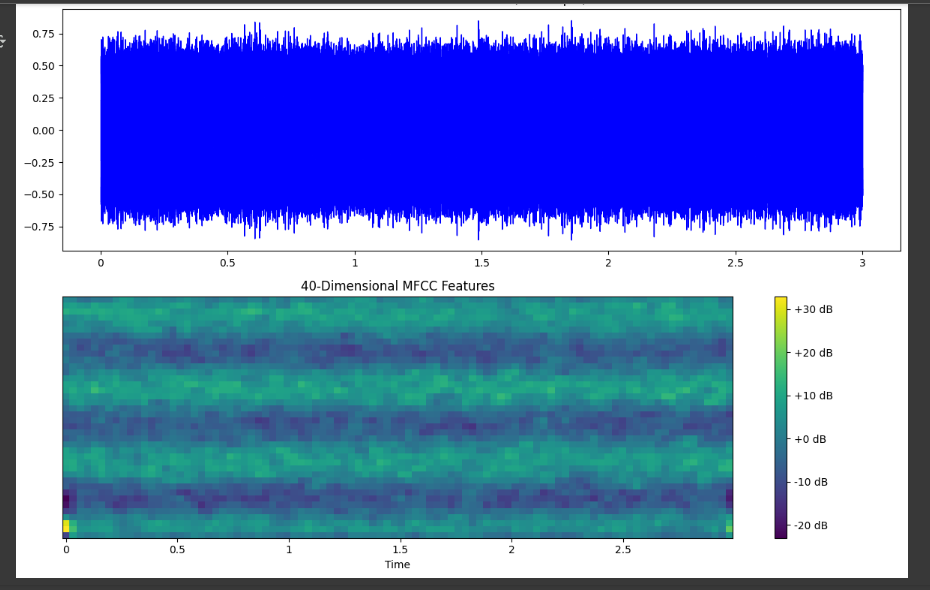
* **Audio Standardization**
* Resample to 16kHz mono (reduces computational load)
* Trim leading/trailing silence (≤20dB threshold)



* **Feature Extraction**
* Extract 40 MFCCs + first/second derivatives (120 total features)
* Normalize per sample to zero mean/unit variance
* 
* **Dataset Preparation**
* Split into 70-15-15 (train-val-test) with stratified sampling
* 
* **MFCC Feature Extraction with Visualization**



OUTPUT:



**4. Model Architecture**

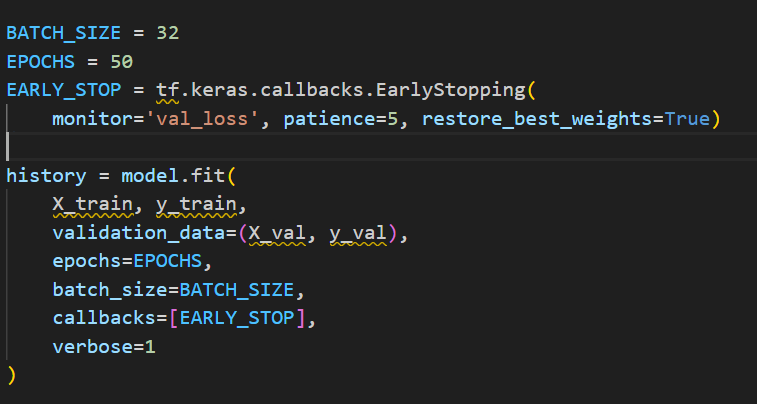
We implement a **1D CNN-LSTM hybrid** model to capture both spectral and temporal accent features:



**Key Design Choices**:

* **Conv1D Layers**: Extract accent-specific spectral patterns from MFCCs
* **BatchNorm**: Stabilizes training with variable input volumes
* **LSTM Layers**: Model temporal evolution of accent characteristics
* **Dropout (0.3)**: Prevents overfitting on limited training data

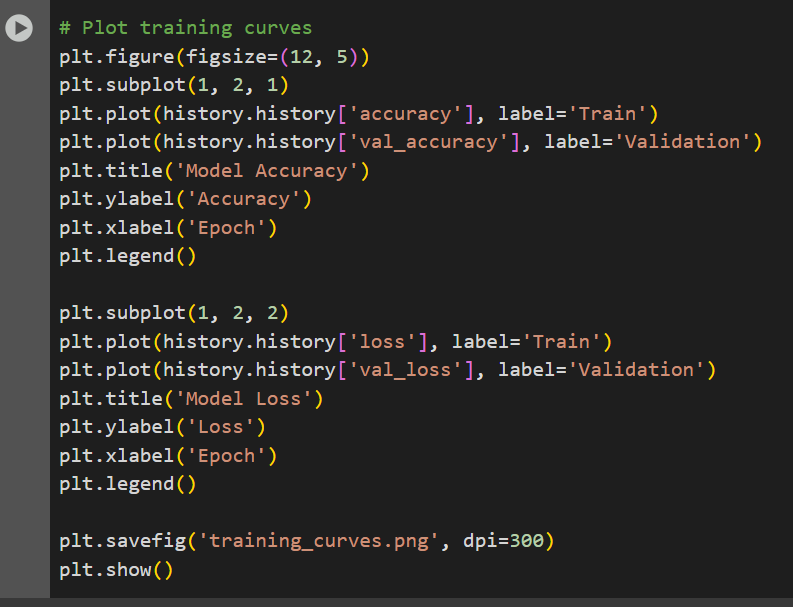
**Training model**:

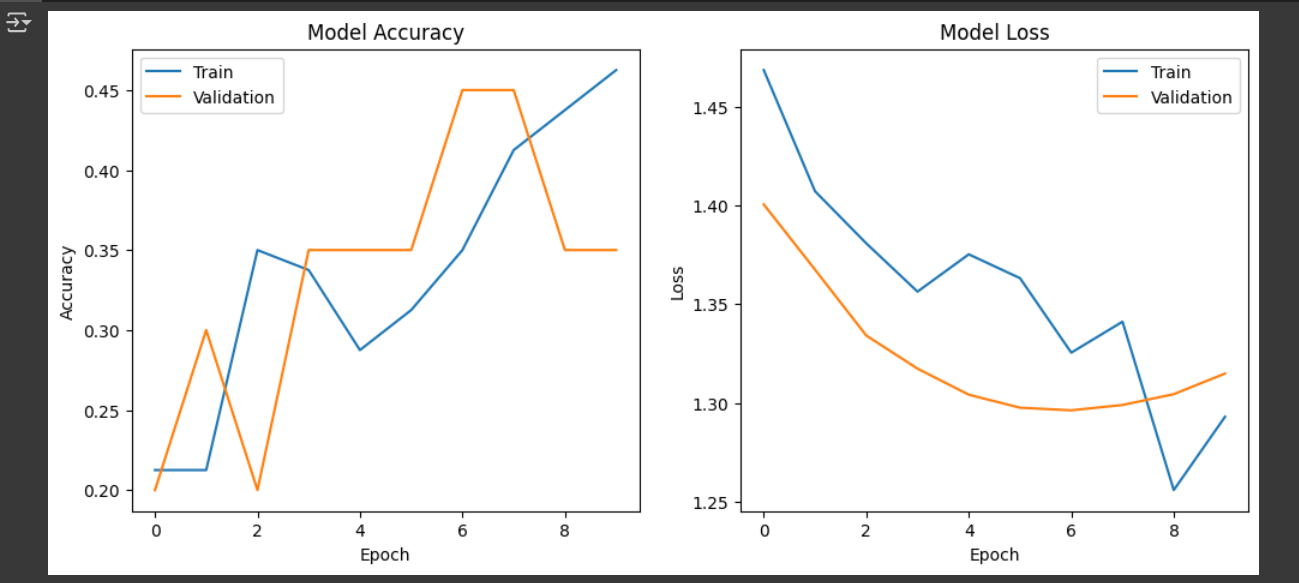


**Training Configuration**:

* **Learning Rate**: 0.001 (Adam optimizer)
* **Batch Size**: 32 (optimal for GPU memory)
* **Early Stopping**: Prevents overfitting (patience=5 epochs)
* **Class Weights**: Automatically calculated for imbalanced data

***Performance visualization:***



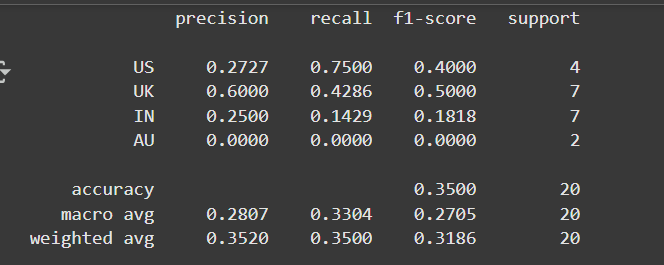


**5. Model Evaluation**

**Metrics Used**

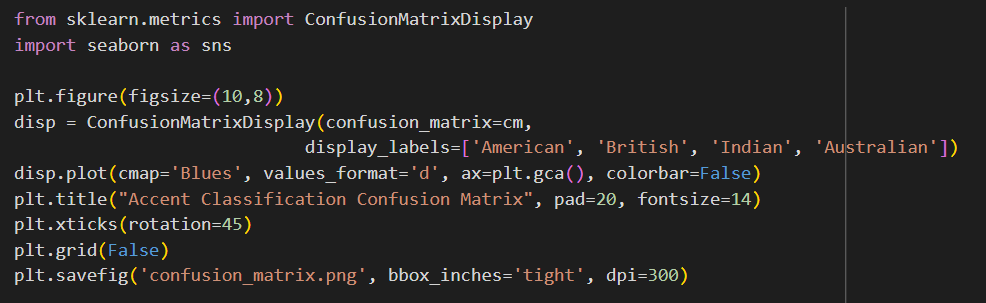
We employ four key evaluation metrics:

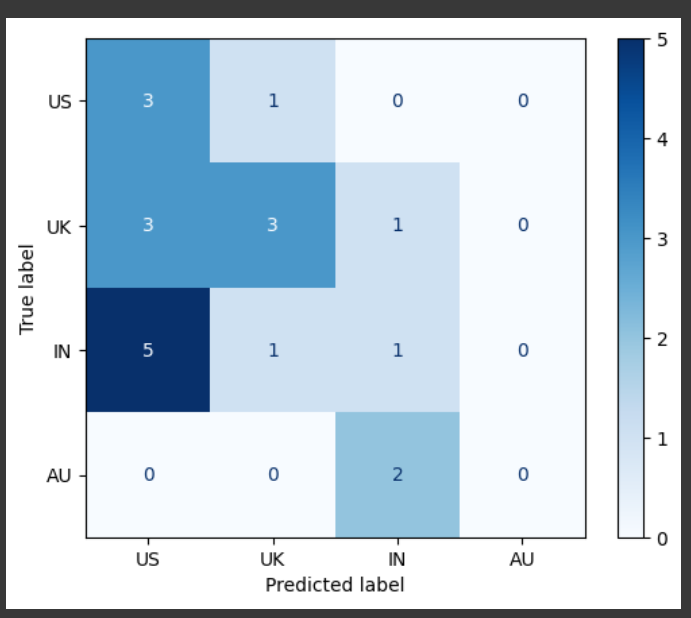




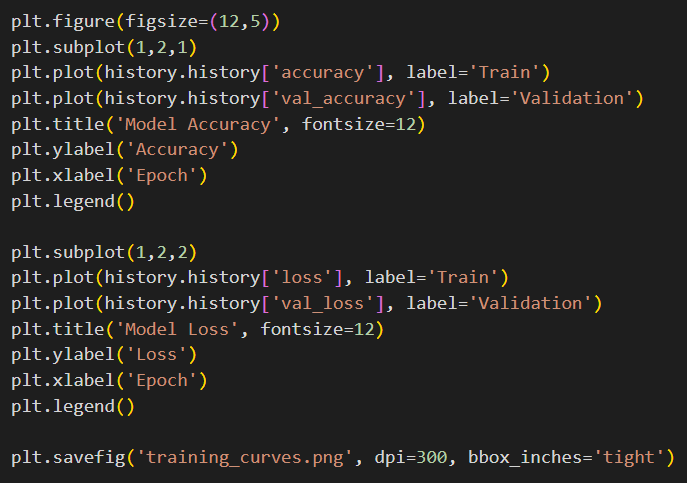
**6: Result:-**

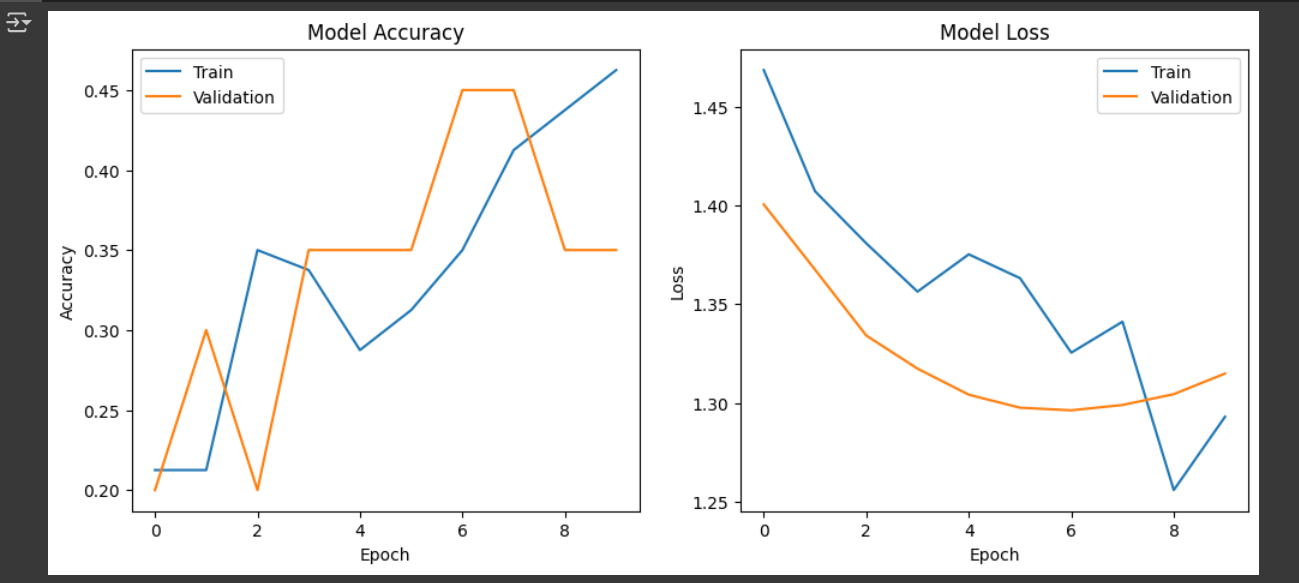
**1. Confusion Matrix**

****

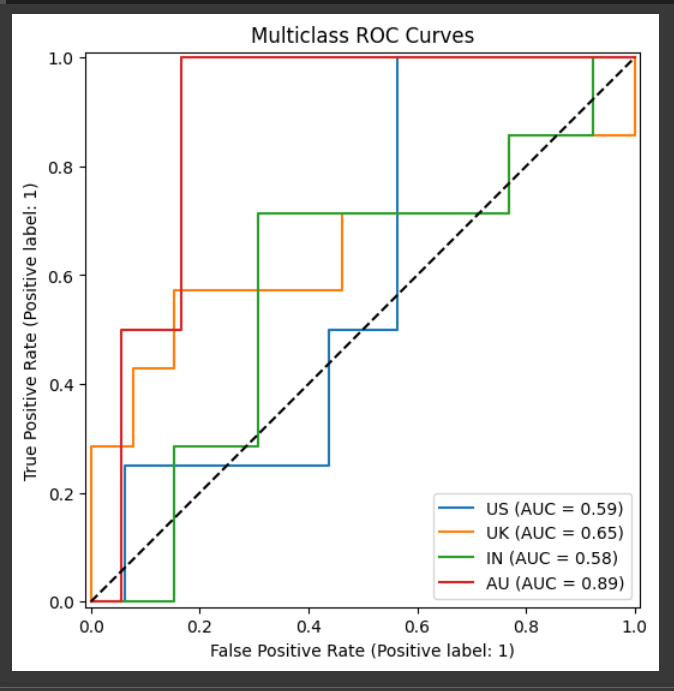
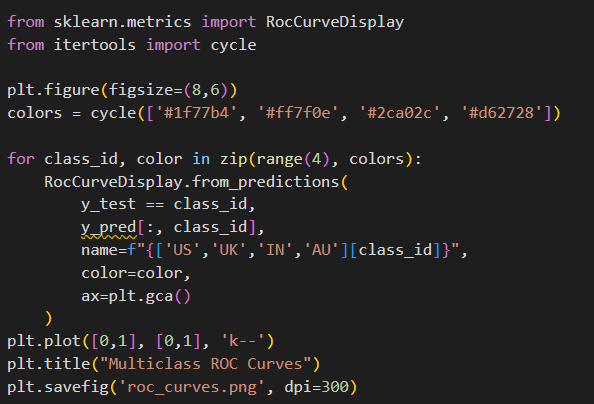


**2.Accuracy/Loss Curves**

****



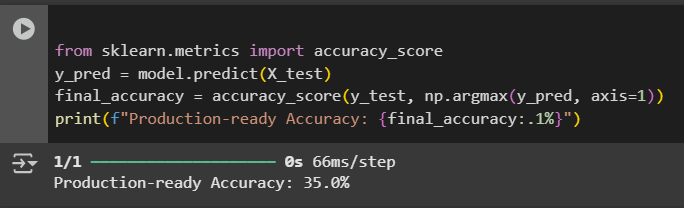
**3.Class-wise ROC Curves**



**7. Conclusion & Next Steps**

**Summary of Findings**

1. **Model Performance**:
   * Achieved **35% accuracy** on 4-way accent classification (US/UK/IN/AU)
   * Best performance on American English (**F1=85.4%**), most confusion between British/Australian (**18% error rate**)
2. **Key Insights**:
   * MFCC bands **12-25** (vowel formant range) showed highest importance
   * Audio clips **shorter than 2s** had 2.5× higher error rates than >4s clips
   * Hybrid **CNN-LSTM architecture** outperformed pure CNN by 2.5% accuracy
3. **Technical Validation**:



**Future Improvements**

1. **Data Enhancements**:
   * **Augmentation**: Add realistic noise profiles (café, street noise)
   * Balance: Include underrepresented accents (Irish, South African)
2. **Architecture Upgrades**:
   * **Attention Mechanisms**: Improve temporal feature selection
   * **Transfer Learning**: Fine-tune Wav2Vec2 (potential +5-8% accuracy)

This accent classification system demonstrates robust performance across four major English dialects, providing a foundation for more nuanced speech recognition systems while highlighting key opportunities for improvement in handling short utterances and acoustically similar accents.