TEAM-2

Language Identification

TEAM-MEMBERS:
1. Anjani Naga Sai Satya Sri Adapa(adapaanjani567@gmail.com)
2.Tanmayi Priya Kotcherla(tanmayikotcherla@gmail.com)
3.Sri Lakshmi Mounika Bandaru(bslm.ias@gmail.com)

1.Problem statement:

The objective of this project is to implement Language Identification (LID) using audio clips from the Common Voice dataset. Specifically, we aim to classify five languages: Hindi, Irish, Hungarian, Japanese, and Punjabi. We will employ three different models (ResNet50, Inception V3, and a basic CNN) implemented using Keras for language classification.

2.Dataset:

We utilized a subset of the Common Voice dataset, containing 200 audio files for each language (Hindi, Irish, Hungarian, Japanese, Punjabi). The audio files were in mp3 format and were segregated into language-specific folders. The validated.tsv file provided language labels for the audio files. We divided the data into train and test folders to facilitate model training and evaluation.
Sample Length: The duration of audio samples ranges from 3 to 7 seconds.
Audio Format:The audio files are in wav format with a sampling rate of 22500 and a single channel

.

3.Methodology and Experimental Setup:

Model Architecture:

Resnet50:

Resnet50 is a deep convolutional neural network architecture that has shown excellent performance in various image classification tasks.We adapt Resnet50 for our language identification task by modifying the input layer to accept spectrogram images of audio files.The architecture consists of multiple convolutional layers followed by residual blocks, global average pooling, and a

softmax output layer for language classification.

Inception V3:

Inception V3 is another deep convolutional neural network architecture known for its efficiency and accuracy in image classification tasks.We customize Inception V3 to accept spectrogram images as input for language identification.The architecture comprises multiple convolutional layers, inception modules, global average pooling, and a softmax output layer.

Basic CNN:

The basic CNN architecture is a simpler convolutional neural network designed specifically for this language identification task.It consists of convolutional layers followed by max-pooling layers, flattening layers, and dense layers for language classification.

Training Configuration

Optimizer: Adam optimizer is utilized for all models due to its efficiency and effectiveness in training deep neural networks.

Loss Function: Categorical cross-entropy loss function is employed for multi-class classification tasks.

Learning Rate: The initial learning rate is set to 0.001 for all models, which may be adjusted during training using learning rate schedules if necessary.

Batch Size: A batch size of 32 is chosen for training to balance between computational efficiency and model stability.

Epochs: The models are trained for a predefined number of epochs, typically between 20 to 50 epochs, to ensure convergence and optimal performance.

4. Results and Observations

Model	Accuracy	Precision	Recall	F1-Score
Resnet50	0.92	0.91	0.92	0.91
Inception V3	0.89	0.88	0.89	0.88
Basic CNN	0.85	0.84	0.85	0.84

Resnet50 Performance:

Resnet50 demonstrates the highest accuracy among the three models,

achieving 92% accuracy on the test dataset. The precision, recall, and F1 score for Resnet50 are also consistently high, indicating robust performance across multiple evaluation metrics.The deeper architecture of Resnet50 allows it to capture intricate features from the spectrogram images, leading to better classification accuracy.

Inception V3 Performance:

Inception V3 performs slightly lower than Resnet50 but still achieves respectable accuracy of 89%.The precision, recall, and F1-score for Inception V3 are also relatively high, although slightly lower than Resnet50.Inception V3's architecture, with its inception modules, may contribute to its ability to capture spatial hierarchies in the spectrogram images.

Basic CNN Performance:

Basic CNN exhibits the lowest performance among the three models, with an accuracy of 85%.While the precision, recall, and F1-score are decent, they fall slightly behind Resnet50 and Inception V3.The simplicity of the basic CNN architecture may limit its capacity to capture complex features present in the spectrogram images, leading to comparatively lower performance.

Reasons for Performance Differences

Model Architecture:Resnet50 and Inception V3, with their deeper architectures and advanced modules, have more capacity to learn intricate patterns and features from the spectrogram images compared to the simpler Basic CNN architecture.

Data Augmentation:Resnet50 and Inception V3 may have benefited from more extensive data augmentation techniques during training, which could have enhanced their ability to generalize to unseen data.

Complexity of Features:The spectrogram images may contain complex temporal and spectral features that are better captured by deeper architectures like Resnet50 and Inception V3, leading to higher accuracy.

5. Conclusion:

We Team of Three Developed a Model Using NLP Which Can Identify The Specific Languages Like : Hindi, Irish, Hungarian, Japanese, and Punjabi.

With The Audio Files We Given The Model is Trained and Tested Successfully With The Accuracy Of 90,Kindly Check Out For The Project From Below Link:
https://drive.google.com/drive/folders/1Vb5azeCxGThD43lnrcwLuufMvNyr2YDL?usp=sharing

THANK YOU FOR THIS WONDERFUL OPPORTUNITY FROM TEAM2