techolution

{ AI done right }

{ Neural Navigators }

Trigger Word Detection

- Aashish
- Shashi
- Mihir
- Kathyayani
- Snigdha

techolution
{AI done right}

Executive Summary

1.Application Focus:
Web app for touch-free interaction. Single-button activation.

2. Functionality: Identifies three door-related commands: open, close, stop. Provides visual feedback based on recognized command.

3. Edge Device Deployment: Model optimized for low-power devices.We focused on deployment on Raspberry Pi.

4. User-Friendly Design: Simple button-driven interface for accessibility.

techolution
{AI done right}

Solution Overview

FRONTEND:

- The Webpage is built using **ReactJS** which establishes a **real-time WebSocket** connection using the **socket.io-client** library, allowing seamless communication with a server.
- It utilizes the **Web Audio API** to access the user's microphone and initiate audio recording.
- It sends recorded audio chunks to the server via WebSocket for real-time processing.
- The client stores server-emitted responses and reacts to trigger command words like "door open," "door close," and "door stop."
- This MVP, showcasing door actions through animations, is primed for real-time integration with Raspberry Pi for practical applications.

techolution
{AI done right}

Solution Overview

BACKEND:

- **Flask** and **Flask_SocketIO** are used for creating a web server and handling socket communication in **real-time.**
- The server receives audio data from the client as a blob and saves it into an .wav file to load using librosa.
- We extract **MFCC** features from the .wav file.
- The pre-trained model uses the MFCC features extracted from client-sent audio signals and makes a prediction.
- The result is then **emitted back** to the client through web socket connection.

techolution
{AI done right}

Solution Overview

Audio Preprocessing:

- **Short time Fourier transform:**
  - computes STFT of an audio signal to analyze its frequency content over time, providing a time-frequency representation essential for extracting features like Mel spectrograms or Mel-Frequency Cepstral Coefficients (MFCCs) in audio signal processing tasks.

- **MFCC Feature Extraction:**
  - The extract_mfcc_features function is called to extract Mel-Frequency Cepstral Coefficients (MFCC) features from the raw audio. The concatenated array of extracted_mfcc_features are the input to our model.

techolution
{AI done right}

Solution Overview

MODEL: CNN

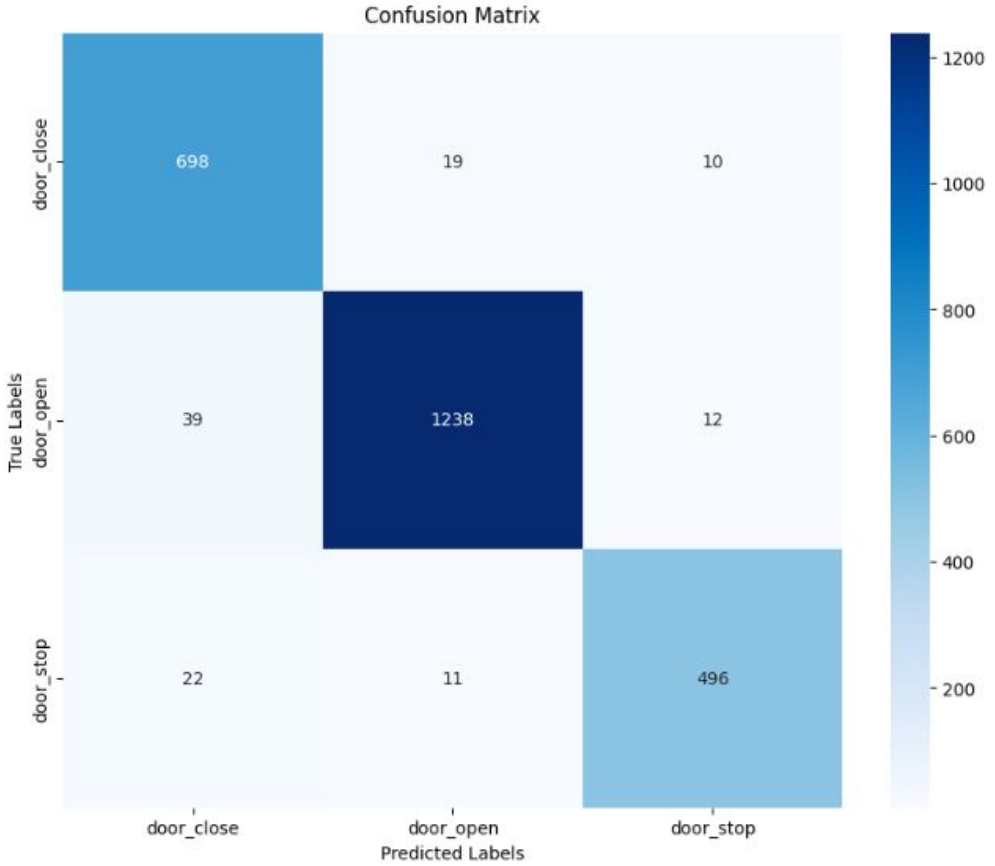
- Two layer CNN model is used here
- **First Conv2D** layer with 8 filters, kernel size (3, 3), and ReLU activation.
- **Second Conv2D** layer with 16 filters, kernel size (3, 3), and ReLU activation.
- **MaxPooling2D** layer with pool size (2, 2) and Dropout layer with a dropout rate of 0.25 for regularization for both layers.
- **Flatten** and **Dense** layer with softmax activation.
- Compilation
  - Adams optimizer
  - Loss function - Sparse categorical crossentropy (suitable for integer-encoded class labels)

techolution
{AI done right}

Solution Overview

MODEL: CNN

- Accuracy - 95.56%
- Precision - 95.62%
- Recall - 95.56%
- F1 Score - 95.57%



techolution
{AI done right}

Solution Overview

MODEL: LSTM

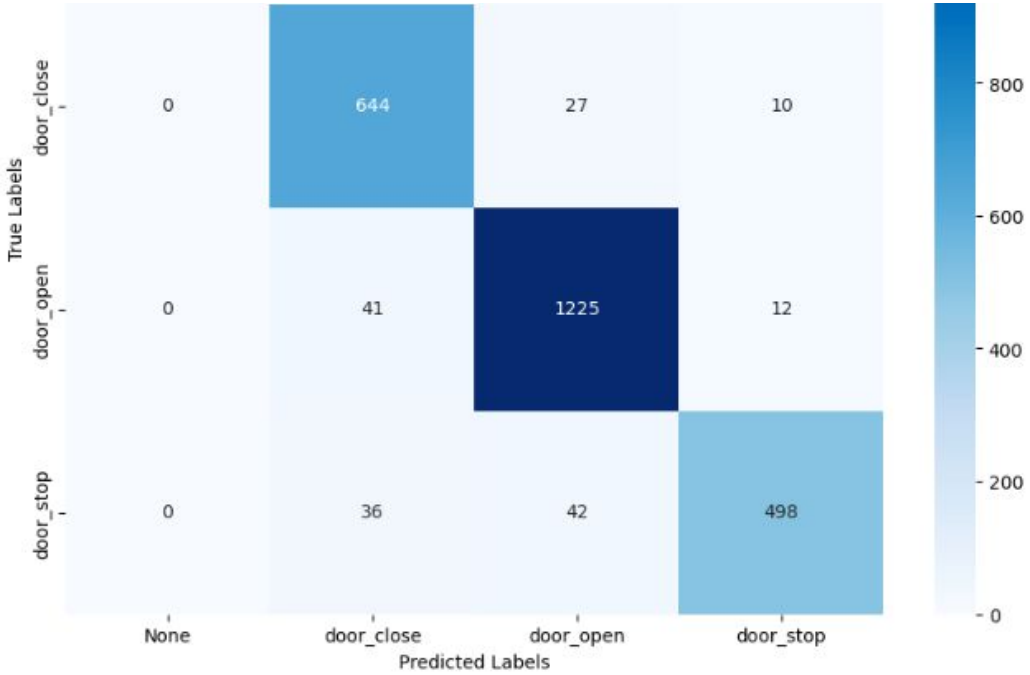
- One layer used
- LSTM layer with 64 units, ReLU activation, and L2 regularization is used for sequence modeling.
- Dropout layer with a dropout rate of 0.2 is added for regularization.
- Dense layer with a sigmoid activation function is used for the output layer with the number of units equal to the number of classes.

techolution
{AI done right}

Solution Overview

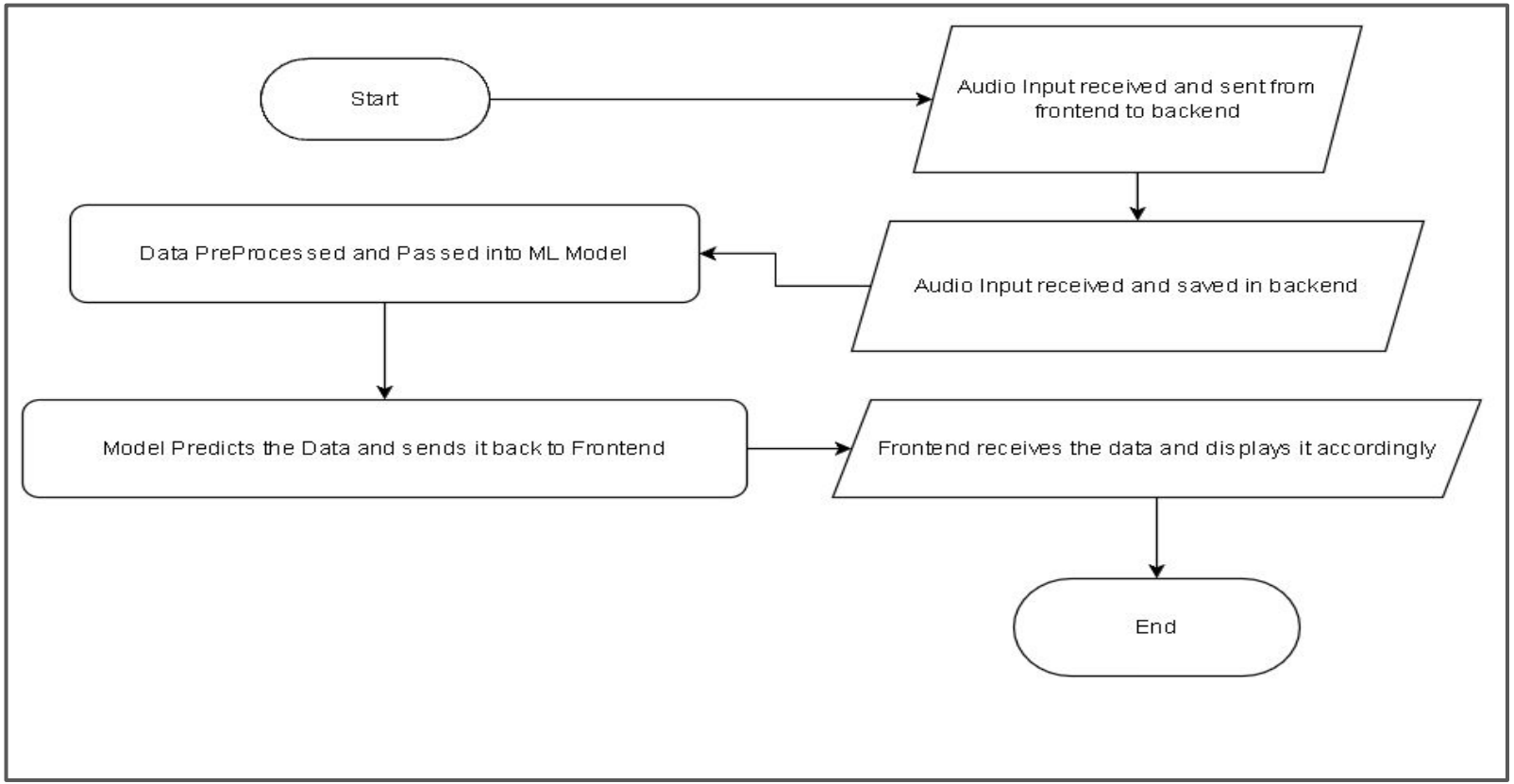
MODEL: LSTM

- Accuracy - 93.37%
- Precision - 93.48%
- Recall - 93.37%
- F1 Score - 93.35%



techolution
{AI done right}

Architecture



Live Demo

In this section you will be providing a live demo of the working of your application.

Raspberry Pi Compatibility

Model Model version: Quantized (int8)

Last training performance (validation set)

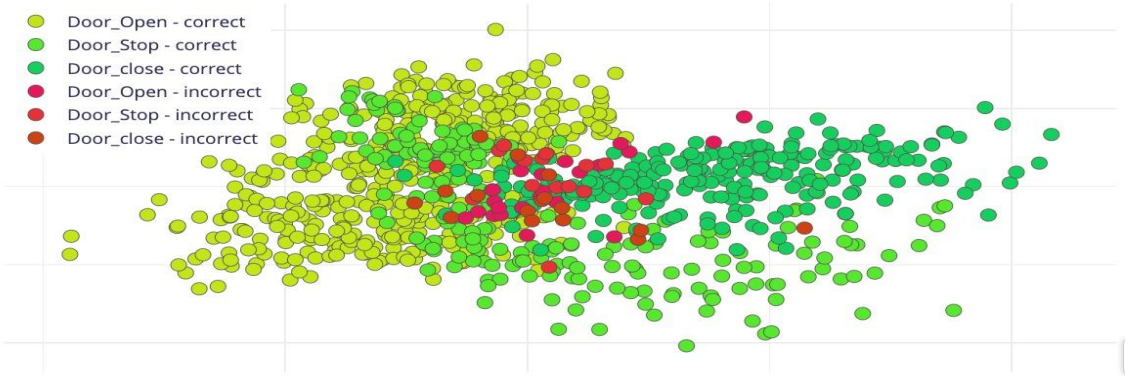
ACCURACY 96.5%

LOSS 0.12

Confusion matrix (validation set)

	DOOR_OPEN	DOOR_STOP	DOOR_CLOSE
DOOR_OPEN	97.3%	0.6%	2.2%
DOOR_STOP	2.8%	95.1%	2.1%
DOOR_CLOSE	2.7%	1.0%	96.3%
F1 SCORE	0.97	0.96	0.95

Data explorer (full training set)



On-device performance

INFERENCING TIME 9 ms.

PEAK RAM USAGE 3.8K

FLASH USAGE 31.4K

techolution
{AI done right}

Conclusion

The versatility of these models renders them applicable across a spectrum of contexts:

- Home Automation:
  - Implementing voice-activated systems enables users to control smart home devices seamlessly without necessitating physical interaction.

- In-Car Voice Control:
  - Integration of the application into automotive systems facilitates safer and more convenient driving experiences, allowing drivers to adjust settings, make calls, or manage entertainment through voice commands.

techolution
Innovation done right

Let's do it right