EE491 - BTP 1:
Voice-controlled interface on RPi Pico W using Zephyr RTOS and TensorFlow (TFLM)

K P Lakshmeesh (22B1246)
EE BTech, IIT Bombay

Guide: Prof. Siddharth Tallur

Why Zephyr?

Built-in Networking stacks

Native support for TFLM

Modular build system

Support for 900+ boards

West, a Zephyr tool - repository management , build + flashing + debugging assist

Some useful modules:
Tflite-micro, CMSIS-NN, CMSIS-DSP, file system, LVGL GUI, Shell interface, Power management, LoRa, OpenThread

RPi-Pico-W

Limited Zephyr support:

- Full support for other Raspberry boards
- Basic functionalities work
- Unsupported: LED, limited Wi-Fi/Bluetooth support

RPi-Pico-W: wifi / bluetooth

```
*** Booting Zephyr OS build v4.2.0-rc3-44-g366d45fd34fe ***
[00:00:03.379,000] ␛[1;33m<wrn> udc_rpi: BUS RESET␛[0m
[00:00:03.468,000] ␛[1;33m<wrn> udc_rpi: BUS RESET␛[0m
[00:00:08.285,000] ␛[0m<inf> MAIN: Turning on AP Mode␛[0m
[00:00:08.287,000] ␛[0m<dbg> net_dhcpv4_server: net_dhcpv4_server_start: Started DHCPv4 server, address pool:␛[0m
[00:00:08.287,000] ␛[0m<dbg> net_dhcpv4_server: net_dhcpv4_server_start:         0: 192.168.4.11␛[0m
[00:00:08.287,000] ␛[0m<dbg> net_dhcpv4_server: net_dhcpv4_server_start:         1: 192.168.4.12␛[0m
[00:00:08.287,000] ␛[0m<dbg> net_dhcpv4_server: net_dhcpv4_server_start:         2: 192.168.4.13␛[0m
[00:00:08.287,000] ␛[0m<dbg> net_dhcpv4_server: net_dhcpv4_server_start:         3: 192.168.4.14␛[0m
[00:00:08.287,000] ␛[0m<inf> MAIN: DHCPv4 server started...
␛[0m
[00:00:08.288,000] ␛[1;33m<wrn> infineon_airoc_wifi: Discard of setting unsupported channel: 255 (will set 1)␛[0m
whd_management_set_event_handler: send event_msgs(iovar) failed
Received buffer request ID: 26853 (expectation: 26854)
Received a response for a different IOCTL - retry
[00:00:13.418,000] ␛[1;31m<err> MAIN: NET_REQUEST_WIFI_AP_ENABLE failed, err: -11␛[0m
[00:00:13.418,000] ␛[0m<inf> MAIN: Connecting to SSID: wifi
␛[0m
[00:00:16.363,000] ␛[0m<inf> MAIN: Connected to wifi␛[0m
[00:00:16.515,000] ␛[0m<inf> net_dhcpv4: Received: 192.168.220.32␛[0m
[00:00:20.556,000] ␛[0m<inf> net_dhcpv4: Received: 192.168.220.32␛[0m
[00:00:28.519,000] ␛[0m<inf> net_dhcpv4: Received: 192.168.220.32␛[0m
```

- Wifi: Limited to STA(station) mode, AP(access point) mode not yet supported
- Bluetooth: Driver implementation incomplete

MAX9814

RPi-Pico-W: ADC

Testing in zephyr:

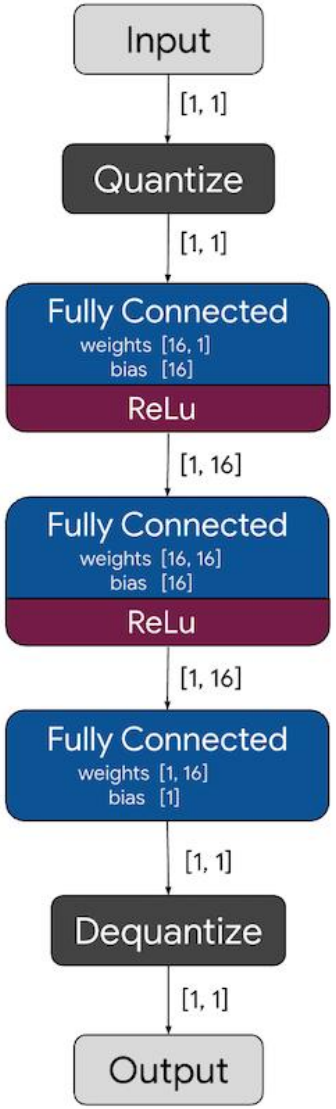
```
/ {
    zephyr,user {
        io-channels = <&adc 0>;
        //io-channels = <&adc 0 &adc 1 &adc 2>; //for multiple channel initialization
    };
};

&adc {
    #address-cells = <1>;
    #size-cells = <0>;

    status="okay";

    channel@0 {
        reg = <0>;
        zephyr,gain = "ADC_GAIN_1";
        zephyr,reference = "ADC_REF_INTERNAL";
        zephyr,acquisition-time = <ADC_ACQ_TIME_DEFAULT>;
        zephyr,resolution = <12>;
    };
    channel@1 {
        reg = <1>;
        zephyr,gain = "ADC_GAIN_1";
        zephyr,reference = "ADC_REF_INTERNAL";
        zephyr,acquisition-time = <ADC_ACQ_TIME_DEFAULT>;
        zephyr,resolution = <12>;
    };
```

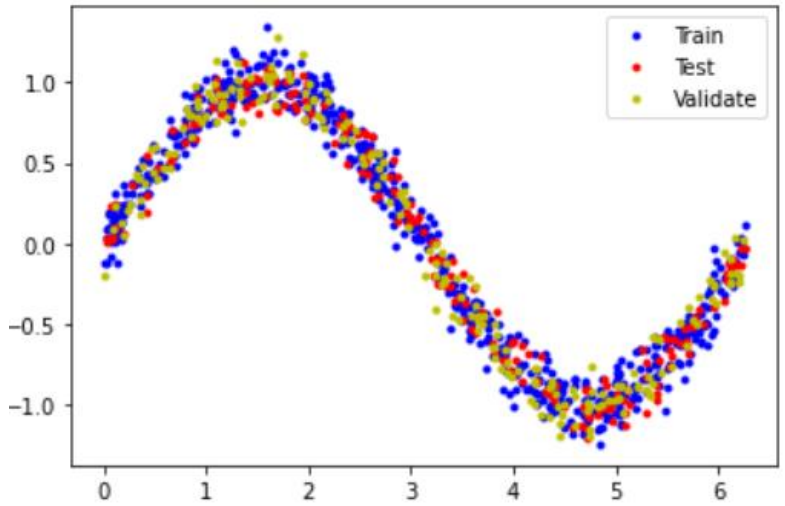
```
ADC reading[0]:
- adc@4004c000, channel 0: 310 = 249 mV
- adc@4004c000, channel 1: 308 = 248 mV
- adc@4004c000, channel 2: 356 = 286 mV
ADC reading[1]:
- adc@4004c000, channel 0: 924 = 744 mV
- adc@4004c000, channel 1: 614 = 494 mV
- adc@4004c000, channel 2: 900 = 725 mV
ADC reading[2]:
- adc@4004c000, channel 0: 900 = 725 mV
- adc@4004c000, channel 1: 599 = 482 mV
- adc@4004c000, channel 2: 915 = 737 mV
ADC reading[3]:
- adc@4004c000, channel 0: 922 = 742 mV
- adc@4004c000, channel 1: 608 = 489 mV
- adc@4004c000, channel 2: 904 = 728 mV
ADC reading[4]:
- adc@4004c000, channel 0: 891 = 717 mV
- adc@4004c000, channel 1: 604 = 486 mV
- adc@4004c000, channel 2: 889 = 716 mV
```

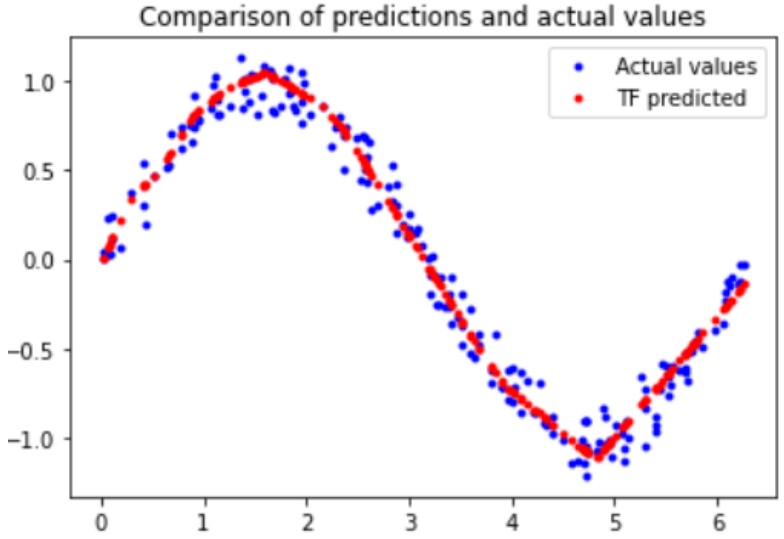
Testing was also done in Micro-Python by converting the ADC samples to wav audio file

- **Limitations:** It takes ~100us to read the ADC, setting a limit of the sampling rate
- **Solution:** shift to 8KHz sampling, Use Direct Memory Access(DMA)

RPi-Pico-W: TFLM

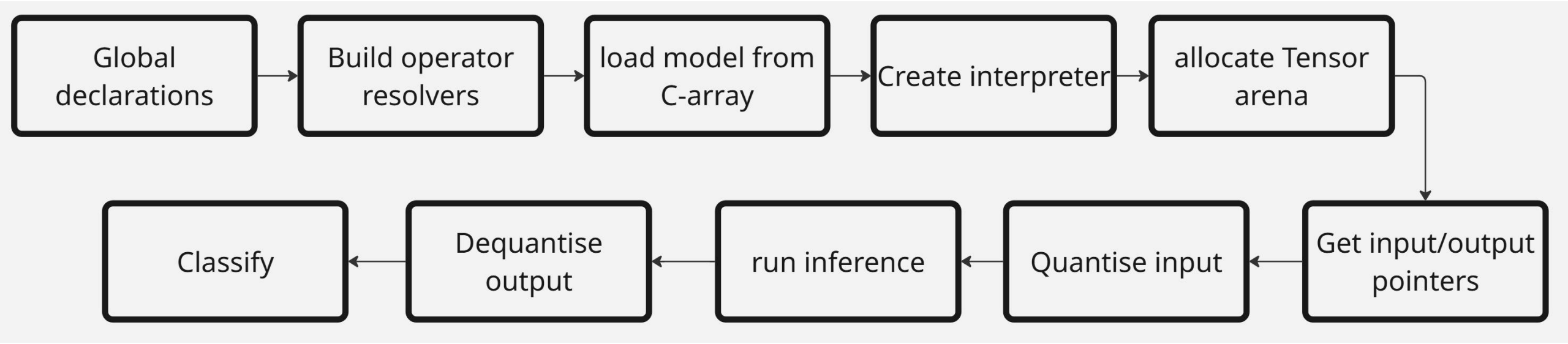






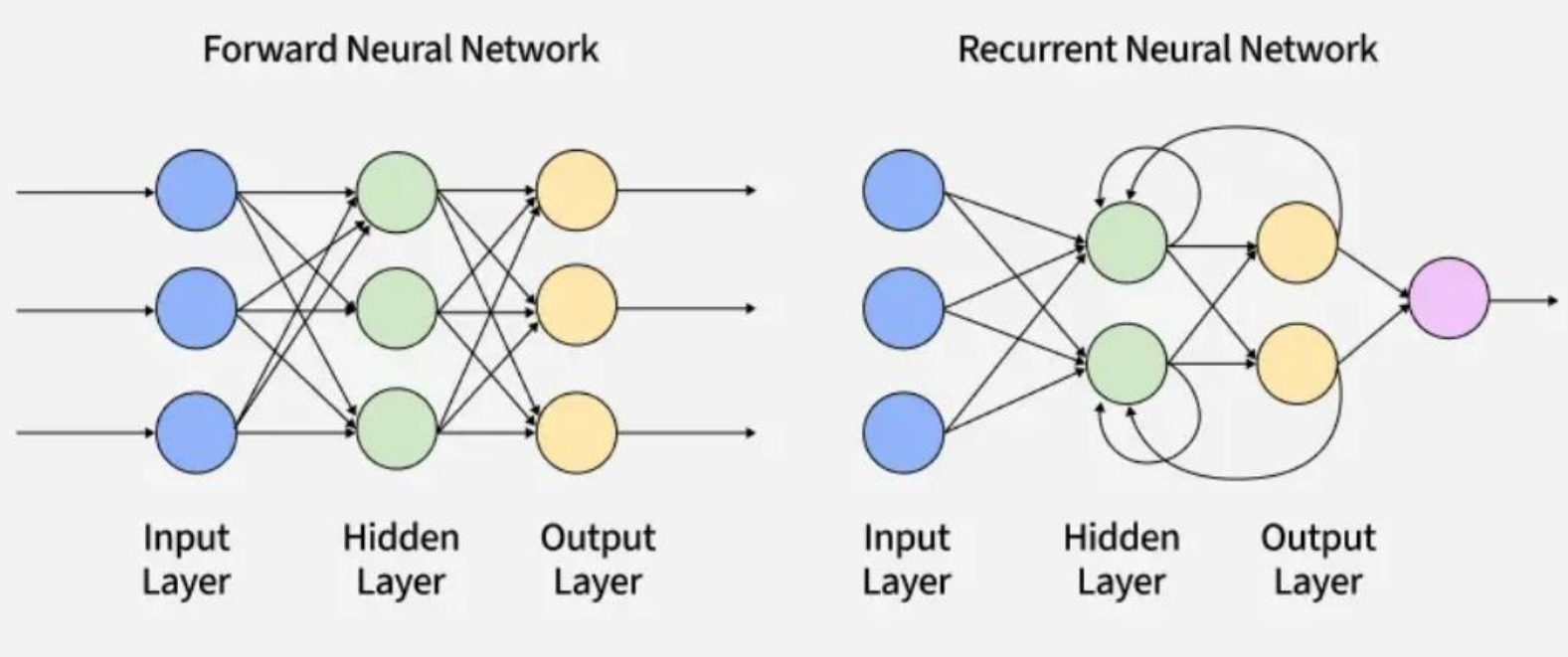

```
x_value: 0.000000, y_value: 0.000000

x_value: 0.314159, y_value: 0.372770

x_value: 0.628319, y_value: 0.559154

x_value: 0.942478, y_value: 0.847203

x_value: 1.256637, y_value: 0.982756
```

TFLM code flow

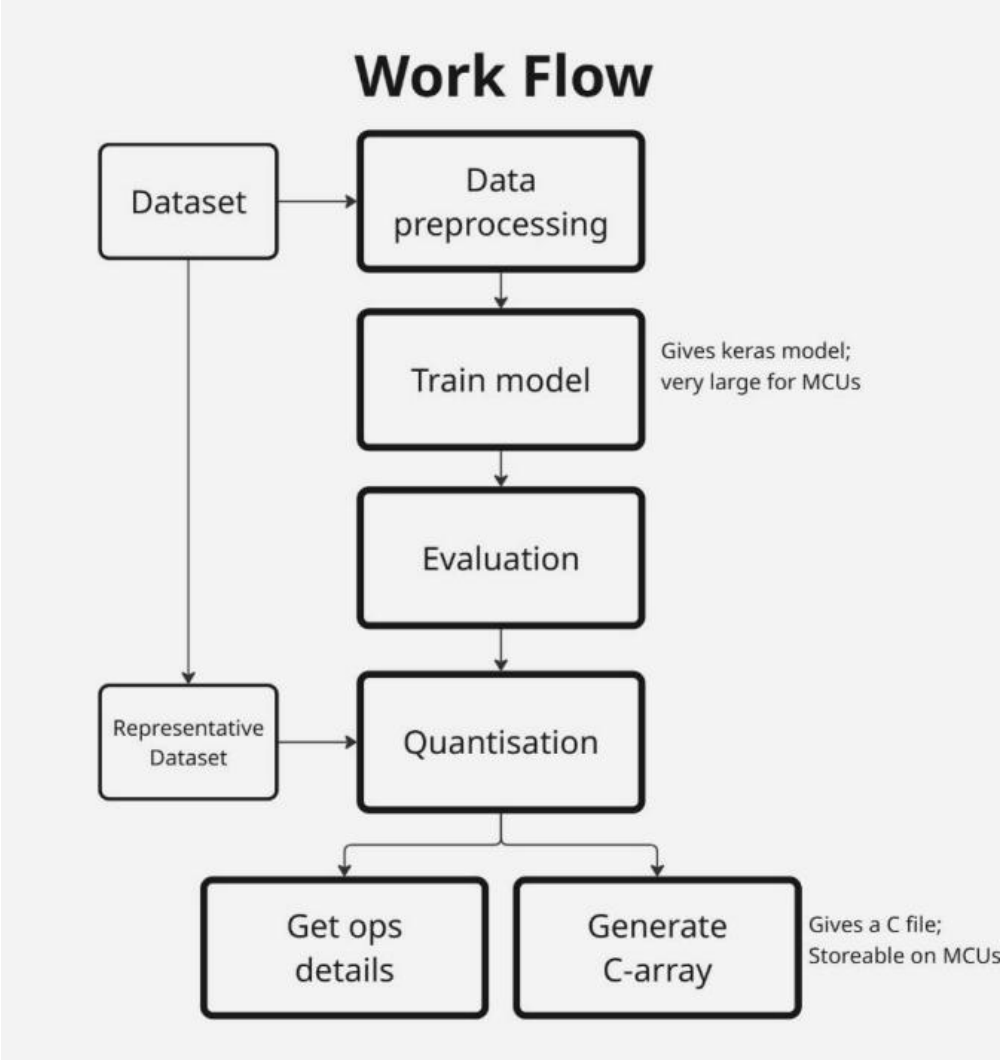


TFLM code flow

- **Limited Ops supported** on TFLM; Dynamic Ops not supported
- **Work-around:** unrolling of model (RAM size may blow-up)



Training the model

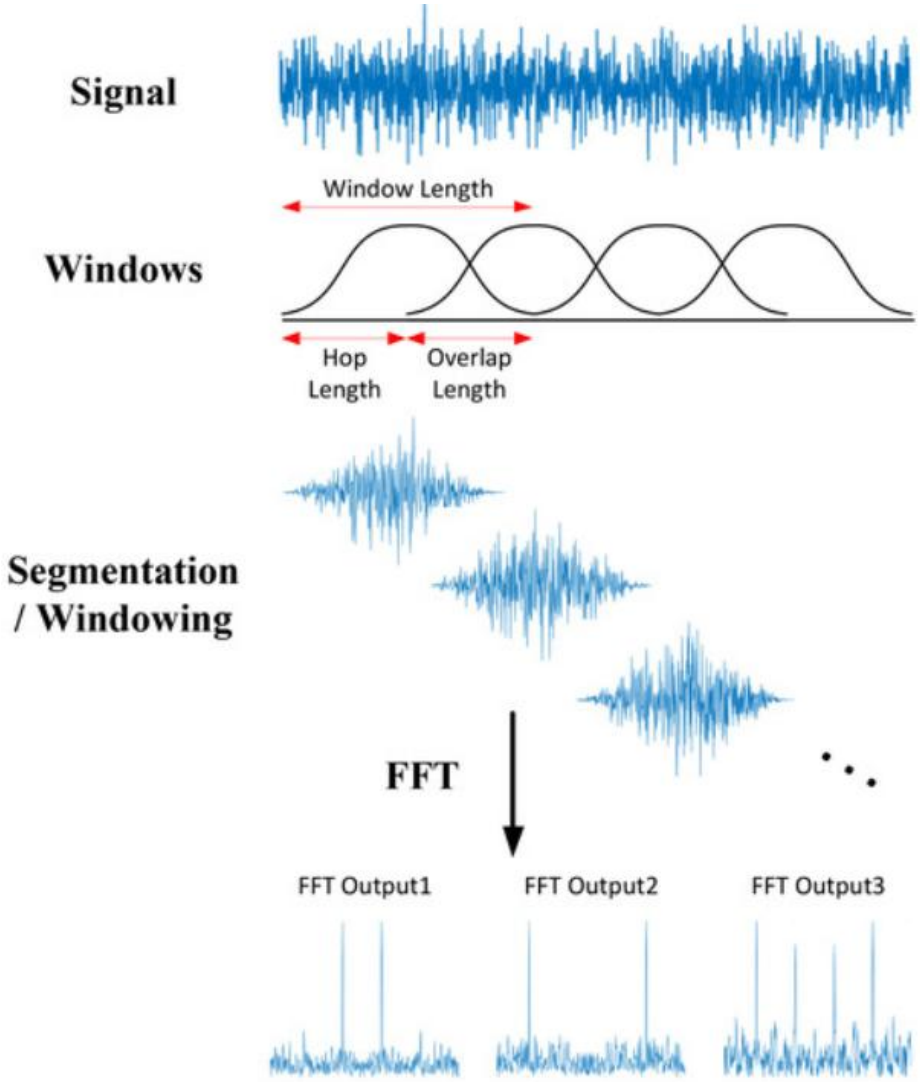


- TensorFlow's keyword recognition example (link)
- Google's dataset for 35 keywords (link)

Model input spectrograms considered:

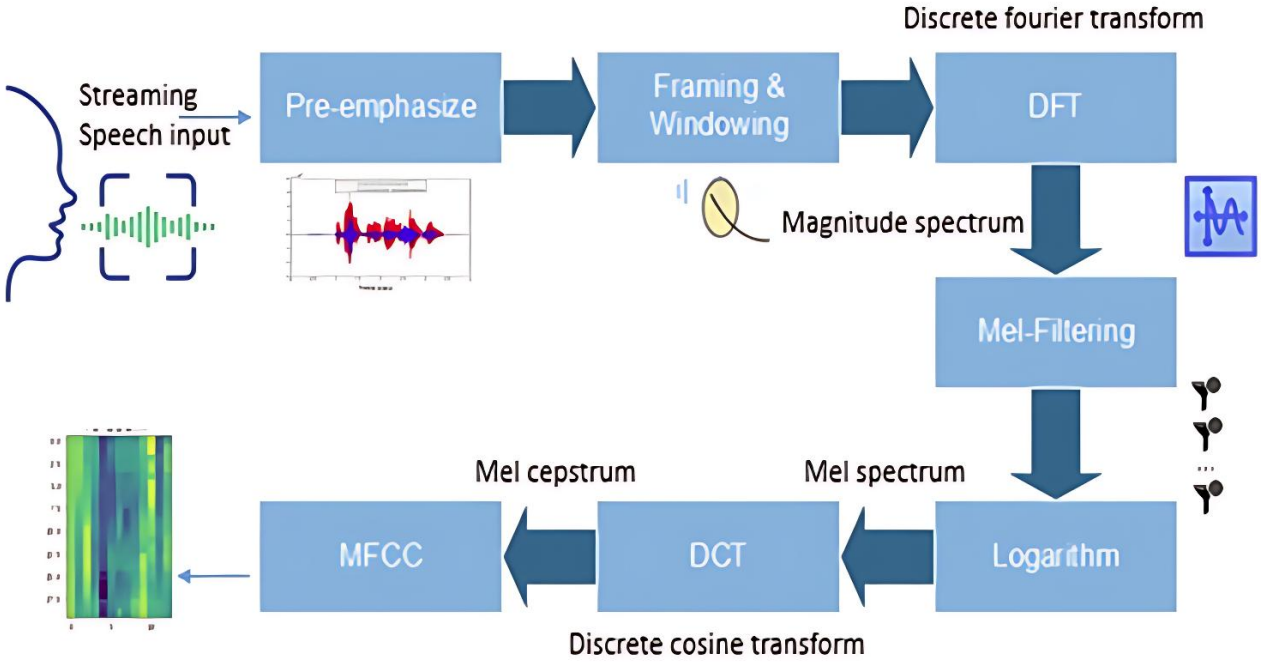
- STFT (short time fourier transform)
- MFCC (Mel-frequency cepstral coefficients)

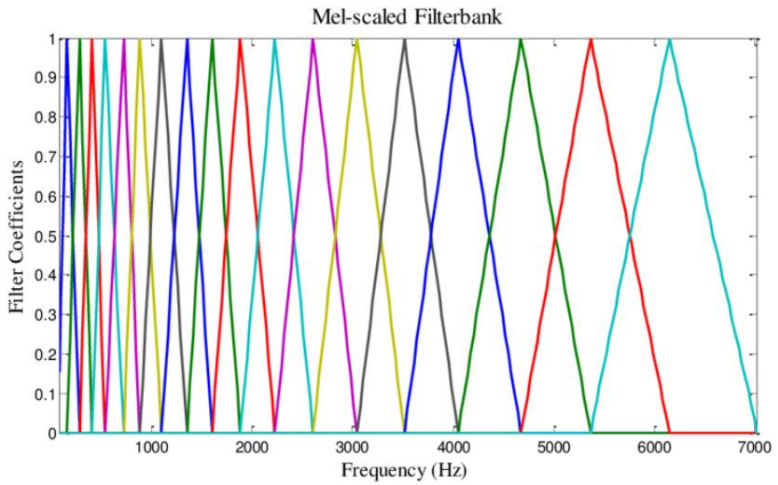
STFT spectrogram



- **Library used:** Kiss-FFT (minimal footprint)
- **Challenge:** Faced issues while building the project

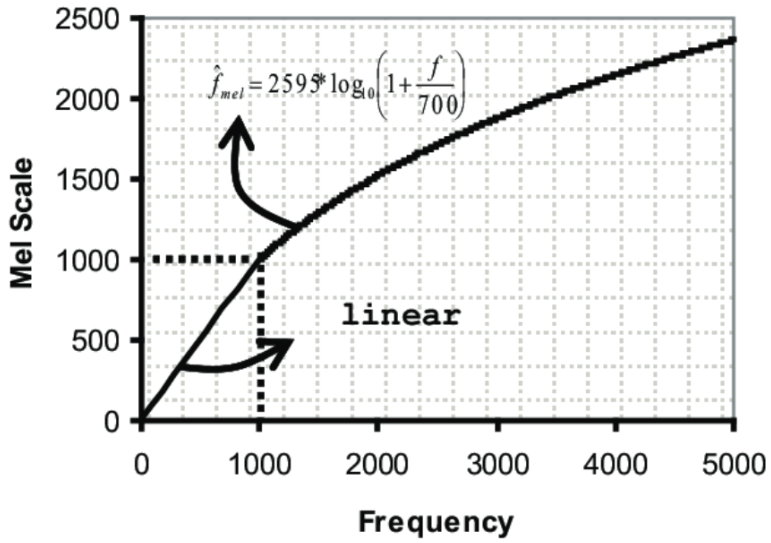
MFCC spectrogram



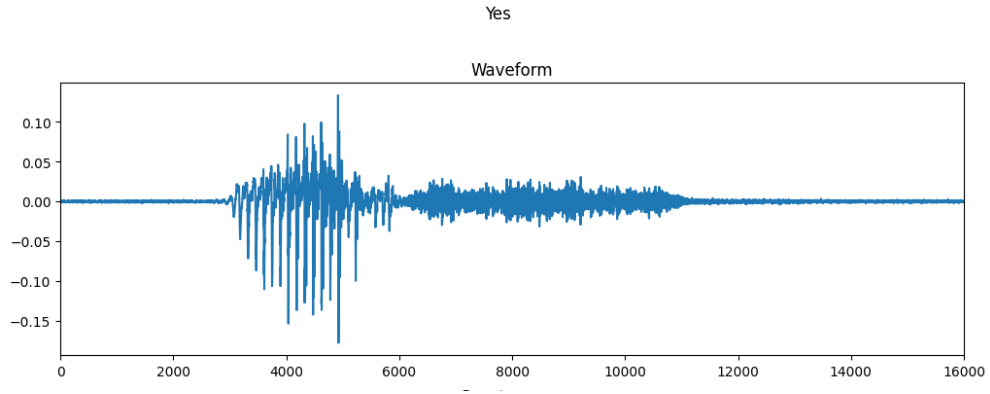



```
dct:
  config1_f32:
    melFilters: 20
    dctOutputs: 13
    type: "f32"

melfilter:
  config1_f32:
    fftlength: 256
    fmin: 64
    fmax: 4000
    samplingRate : 8000
    melFilters: 20
    type: "f32"
window:
  config1_f32:
    fftlength: 256
    type: "f32"
    win: "hamming"
```

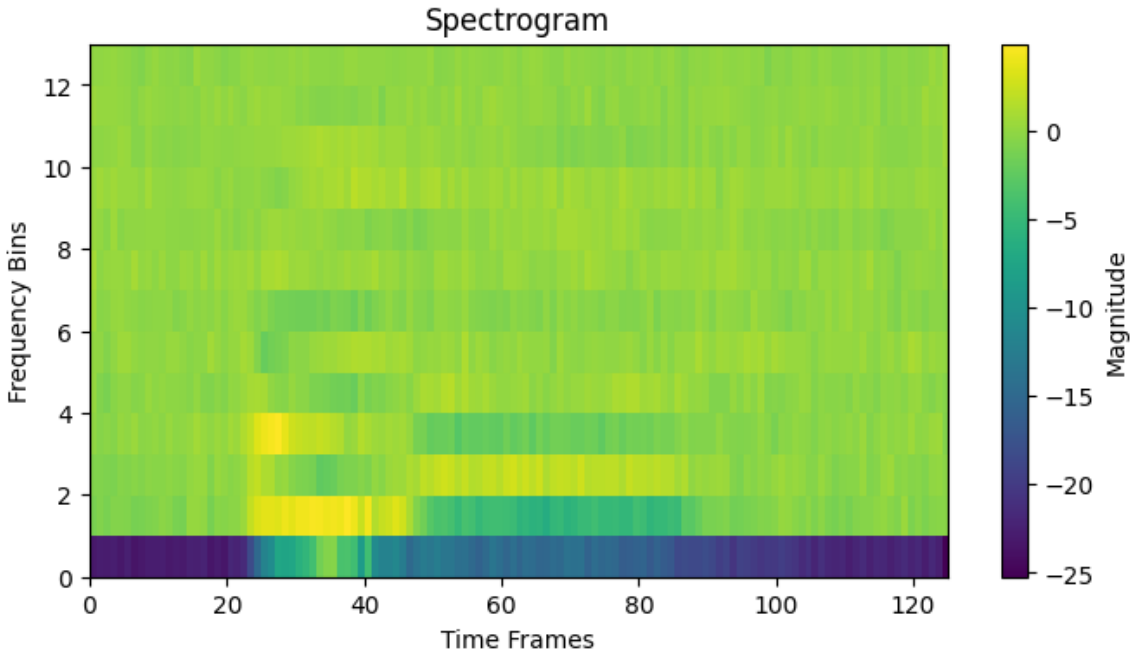
- **Library used:** CMSIS-DSP (optimised for ARM-cortex A/M)
- **Challenge:** loop get halted for unknown reasons.
  Suspected to be memory related issue.



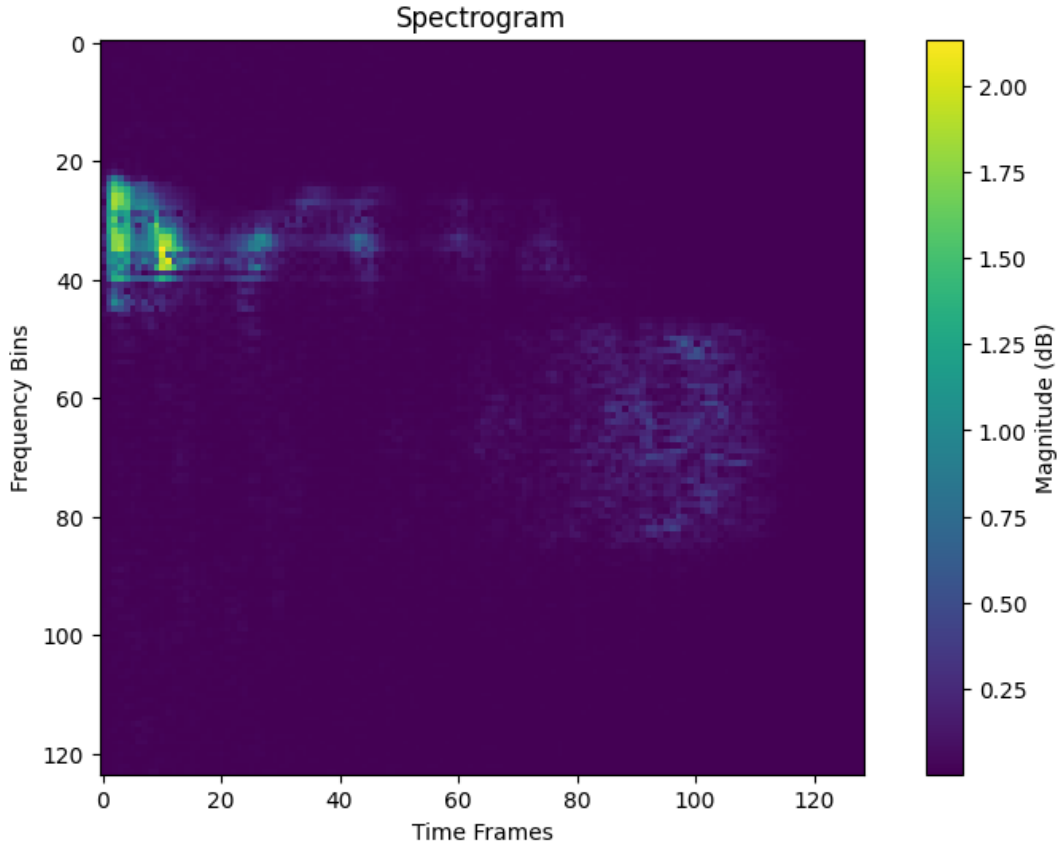
STFT vs MFCC



Dim: 13 x 125



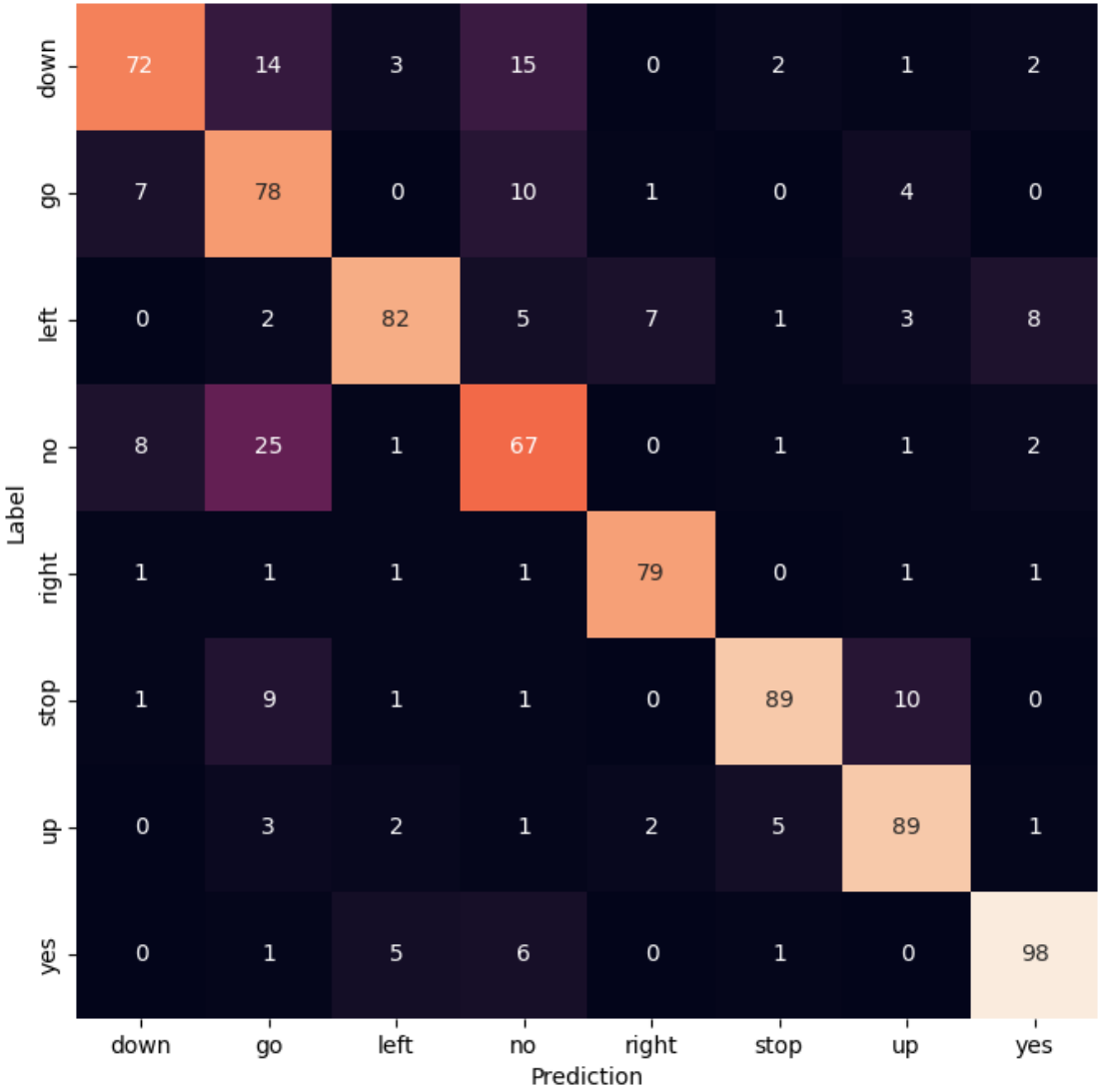
Dim: 124 x 129



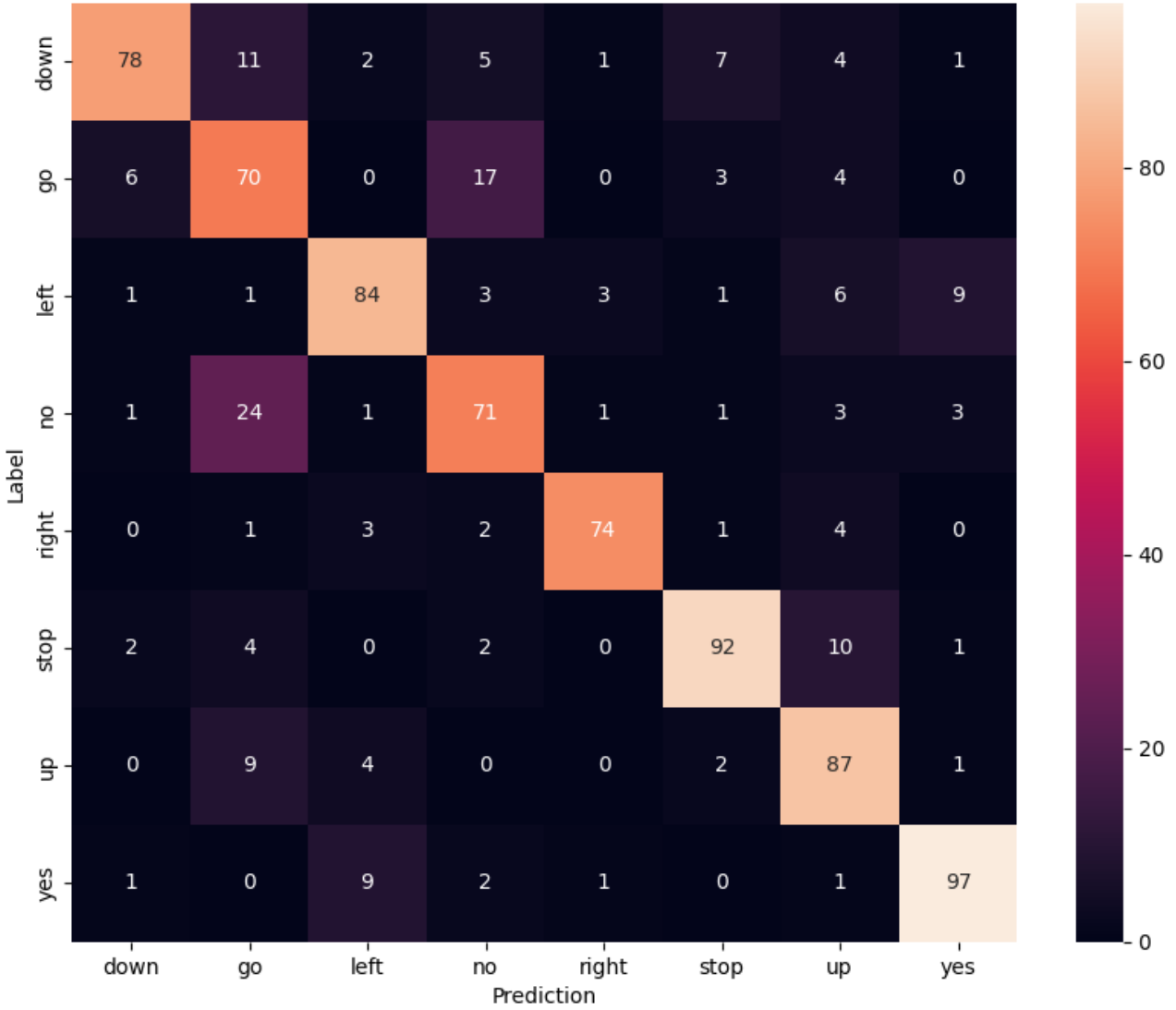
Model	#params	Size	accuracy
STFT input 2D CNN	20K	78KB	77%
MFCC input 1D CNN	15.8K	62KB	82.1

Test results

STFT input 2D CNN:

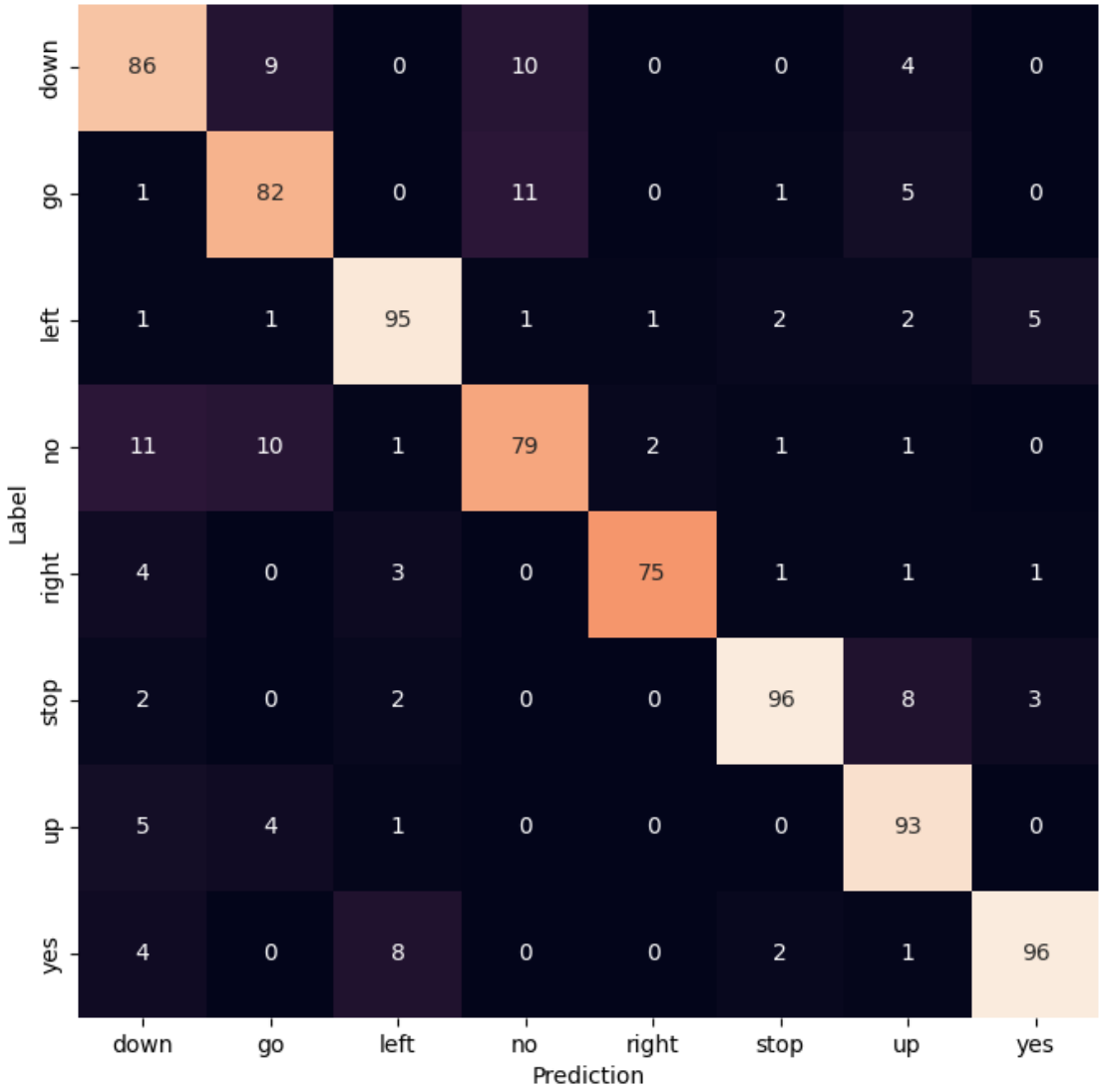


MFCC input 2D CNN:

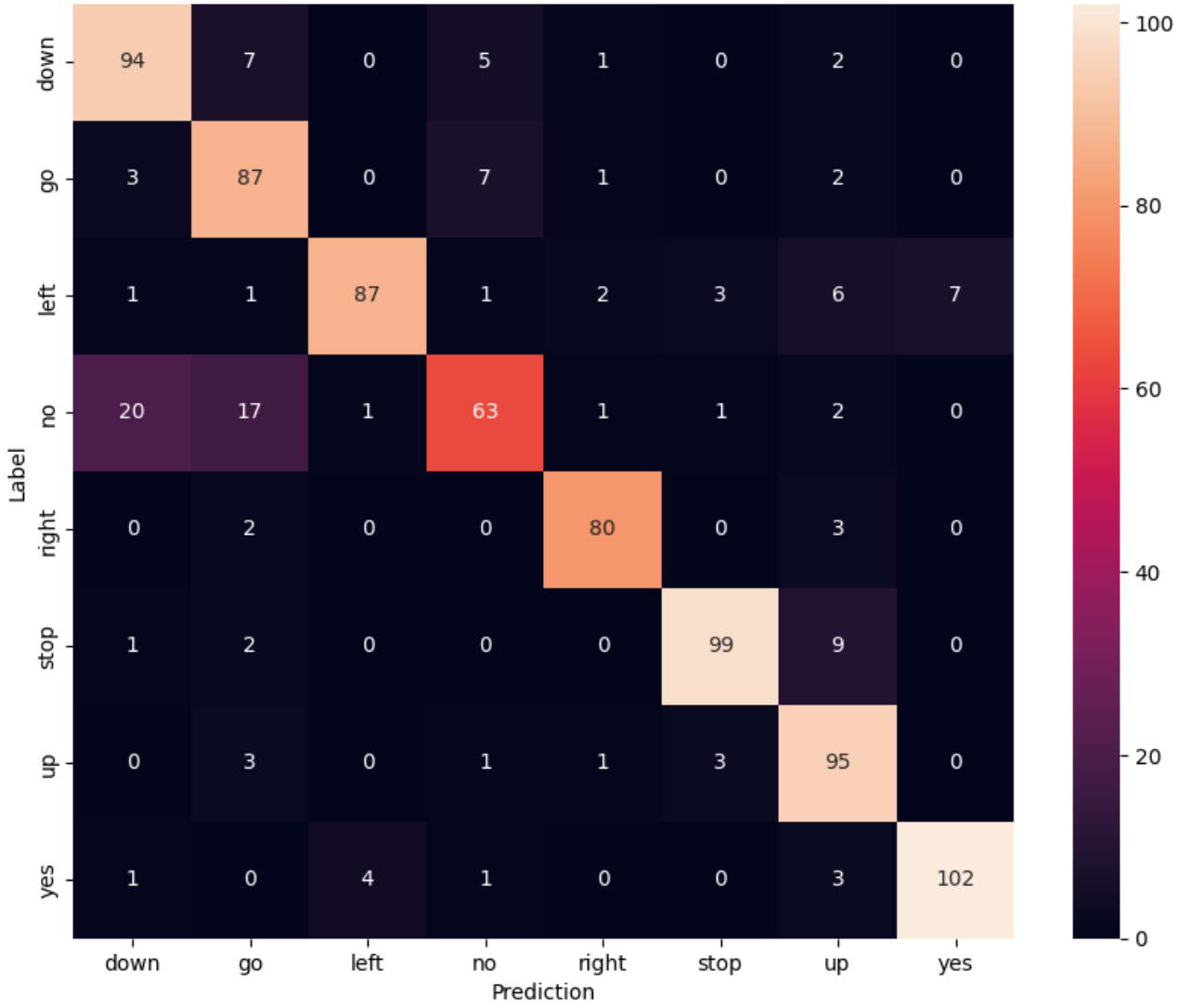


Test results

MFCC input GRU:

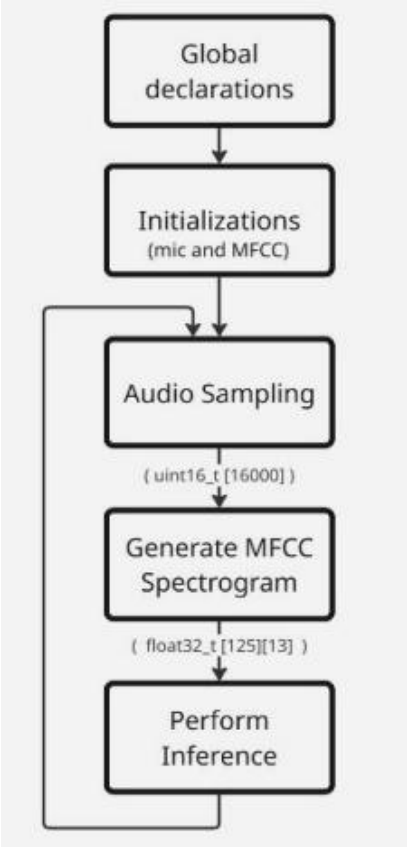


MFCC input LSTM:

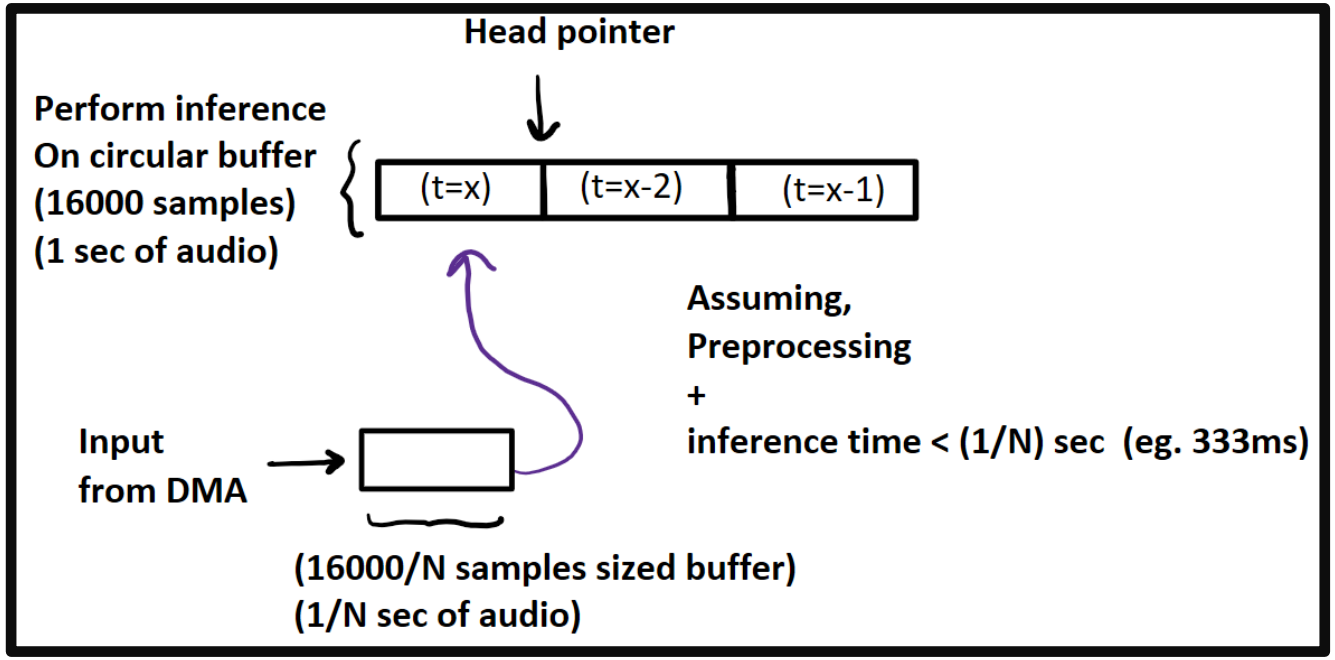


Application flow

Currently:



Future:



Demo testing

CNN Models

Model	Parameters	Size	Accuracy
STFT 2D-CNN	~20K	78 KB	**77.0%**
MFCC 1D-CNN	15.8K	62 KB	**82.1%**

RNN Models

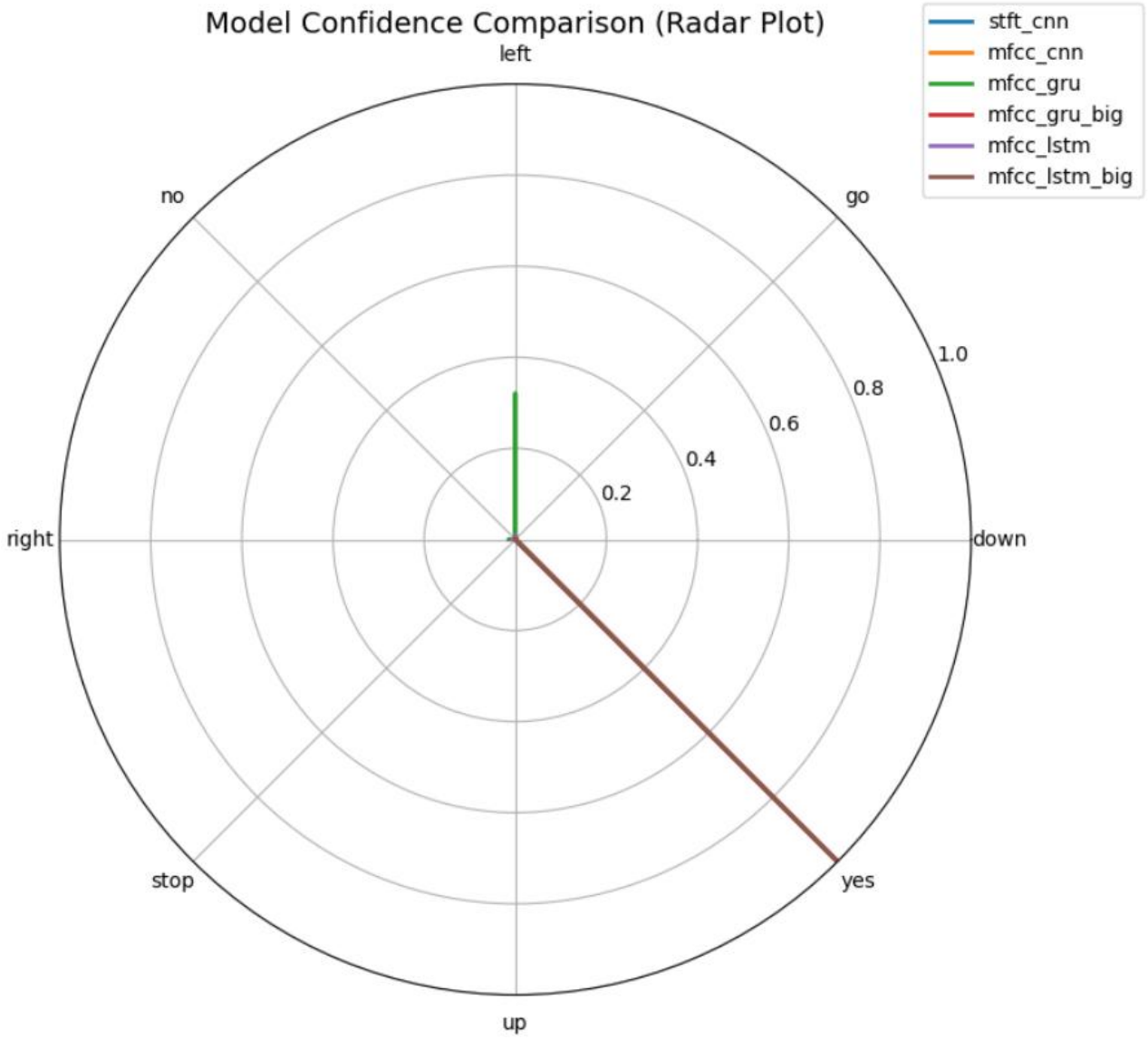
Model	Parameters	Size	Train Time	Accuracy
GRU (Small)	5.8K	22.8 KB	~1:45 min	**85.3%**
GRU (Large)	19.8K	77.5 KB	—	**91.1%**
LSTM (Small)	7.2K	28 KB	~1:00 min	**79.9%**
LSTM (Large)	~14K	55 KB	—	**84.5%**

Inference Confidence

Input: Yes

Inference result:

```
stft_cnn: yes (0.996)
mfcc_cnn: yes (0.996)
mfcc_gru: yes (0.660)
mfcc_gru_big: yes (0.999)
mfcc_lstm: yes (0.980)
mfcc_lstm_big: yes (0.994)
```

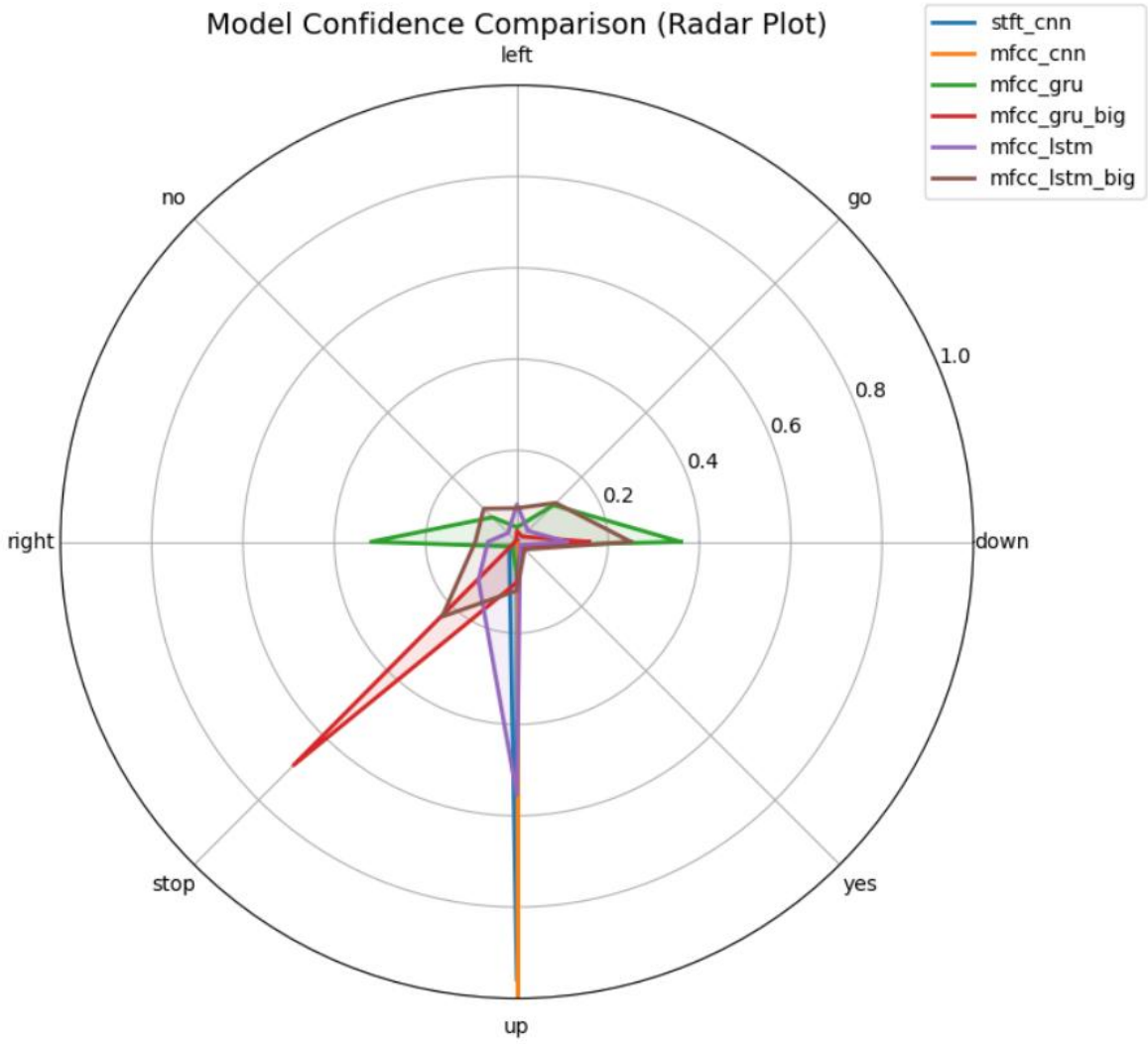


Inference Confidence

Input: up

Inference result:

```
stft_cnn: up (0.961)
mfcc_cnn: up (0.996)
mfcc_gru: down (0.359)
mfcc_gru_big: stop (0.692)
mfcc_lstm: up (0.555)
mfcc_lstm_big: down (0.250)
```

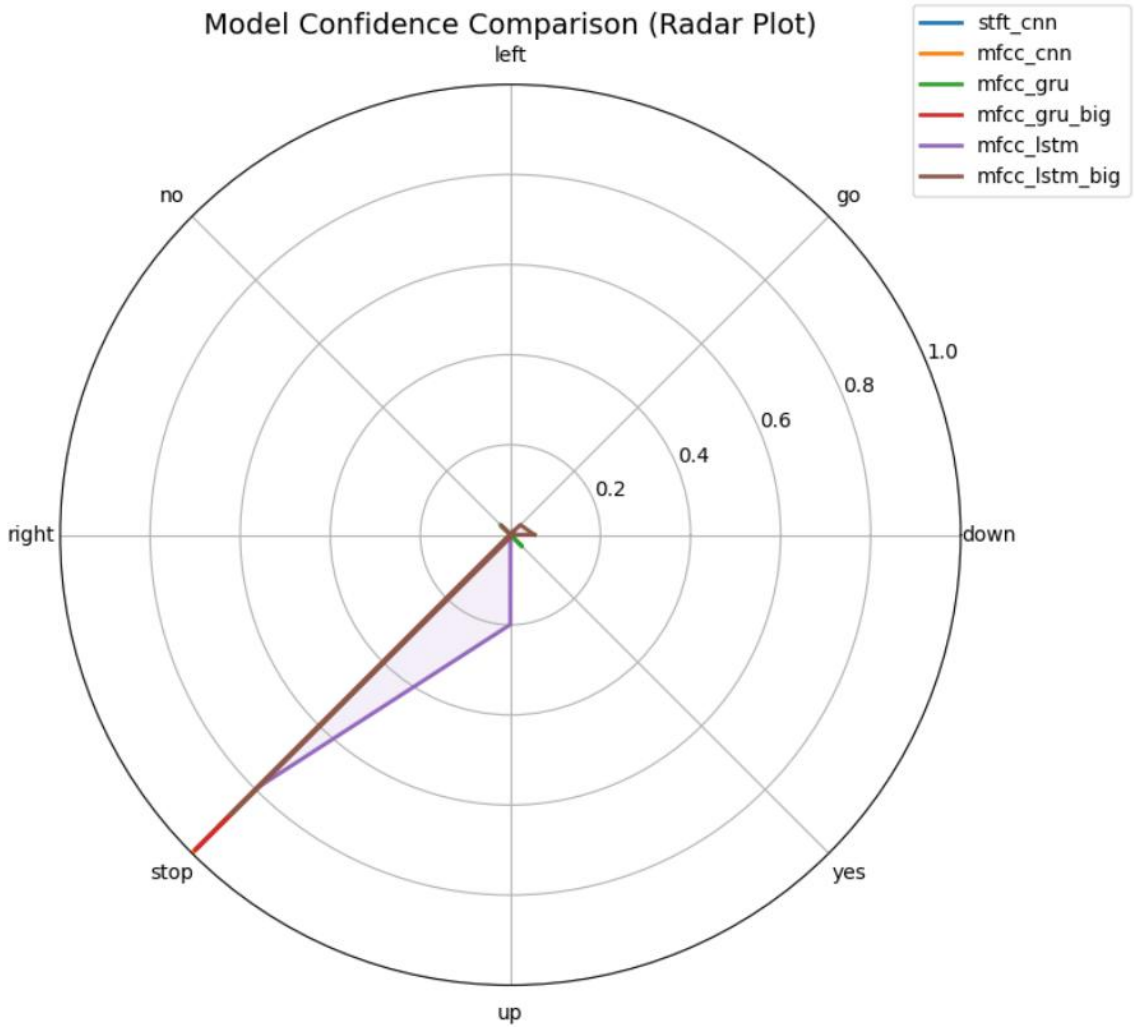


Inference Confidence

Input: stop

Inference result:

```
stft_cnn: stop (0.996)
mfcc_cnn: stop (0.996)
mfcc_gru: stop (0.916)
mfcc_gru_big: stop (0.990)
mfcc_lstm: stop (0.793)
mfcc_lstm_big: stop (0.879)
```

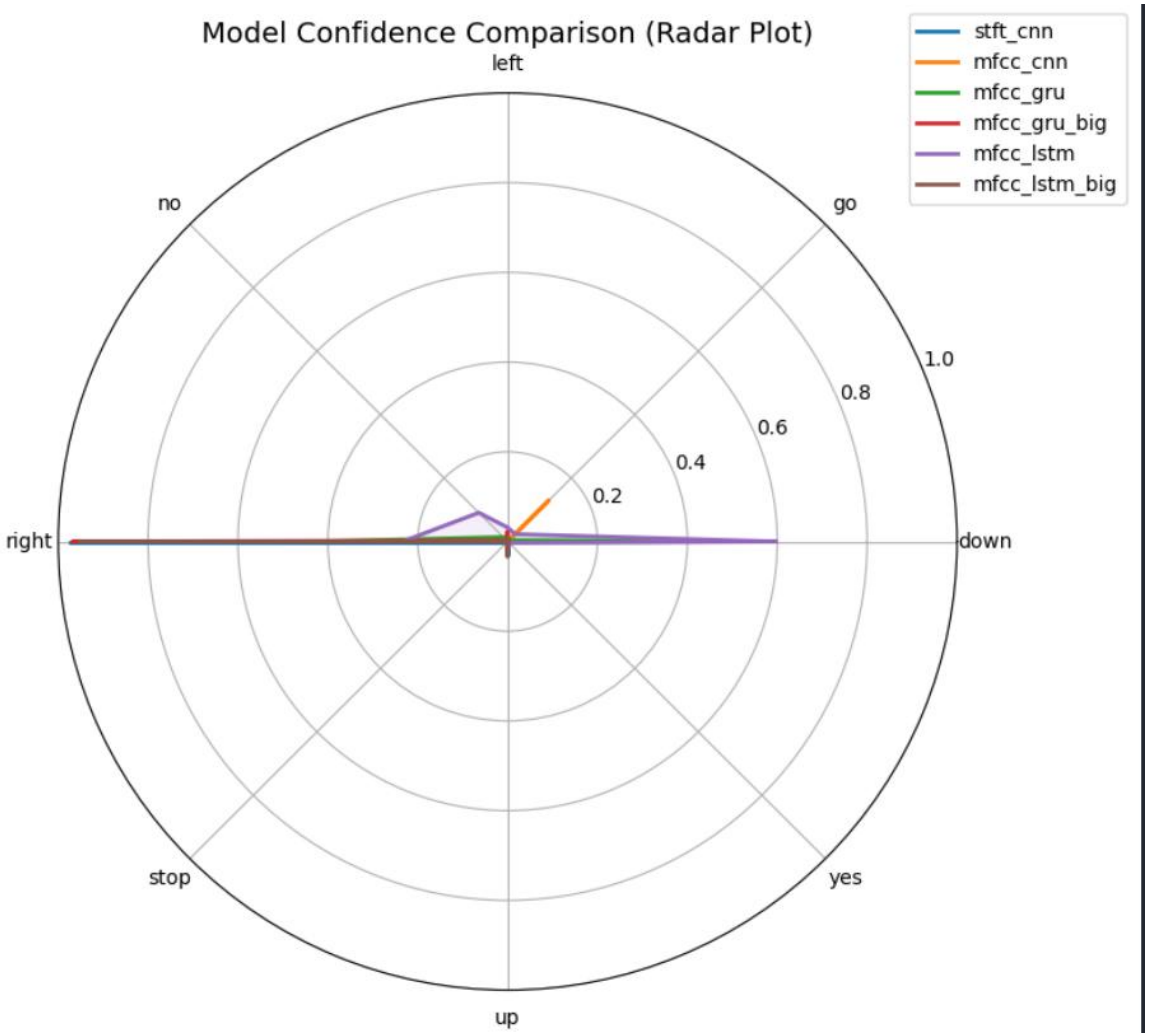


Inference Confidence

Input: right

Inference result:

```
stft_cnn: right (0.973)
mfcc_cnn: right (0.871)
mfcc_gru: down (0.556)
mfcc_gru_big: right (0.968)
mfcc_lstm: down (0.595)
mfcc_lstm_big: right (0.953)
```

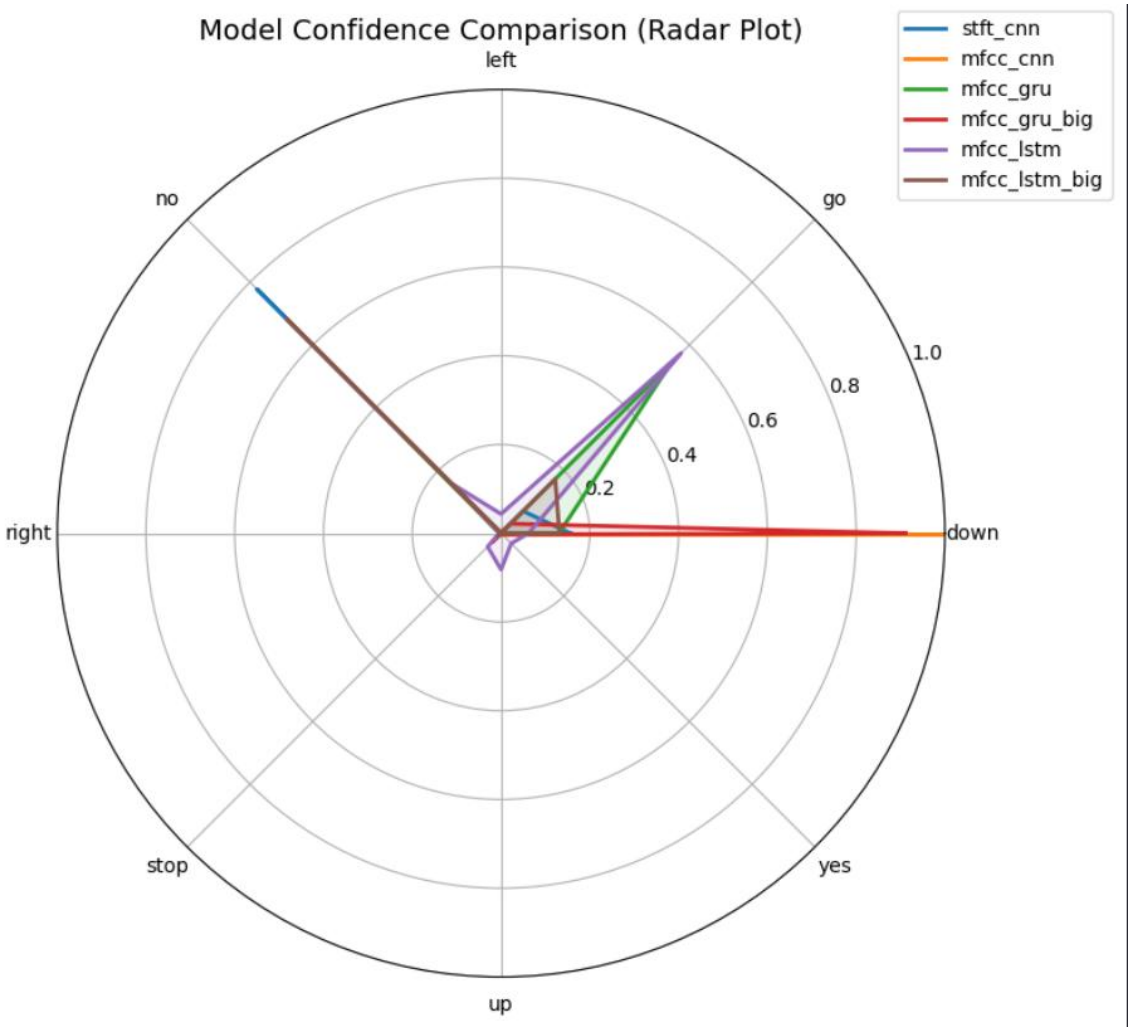


Inference Confidence

Input: no

Inference result:

```
stft_cnn: no (0.777)
mfcc_cnn: down (0.996)
mfcc_gru: go (0.540)
mfcc_gru_big: down (0.911)
mfcc_lstm: go (0.574)
mfcc_lstm_big: no (0.684)
```

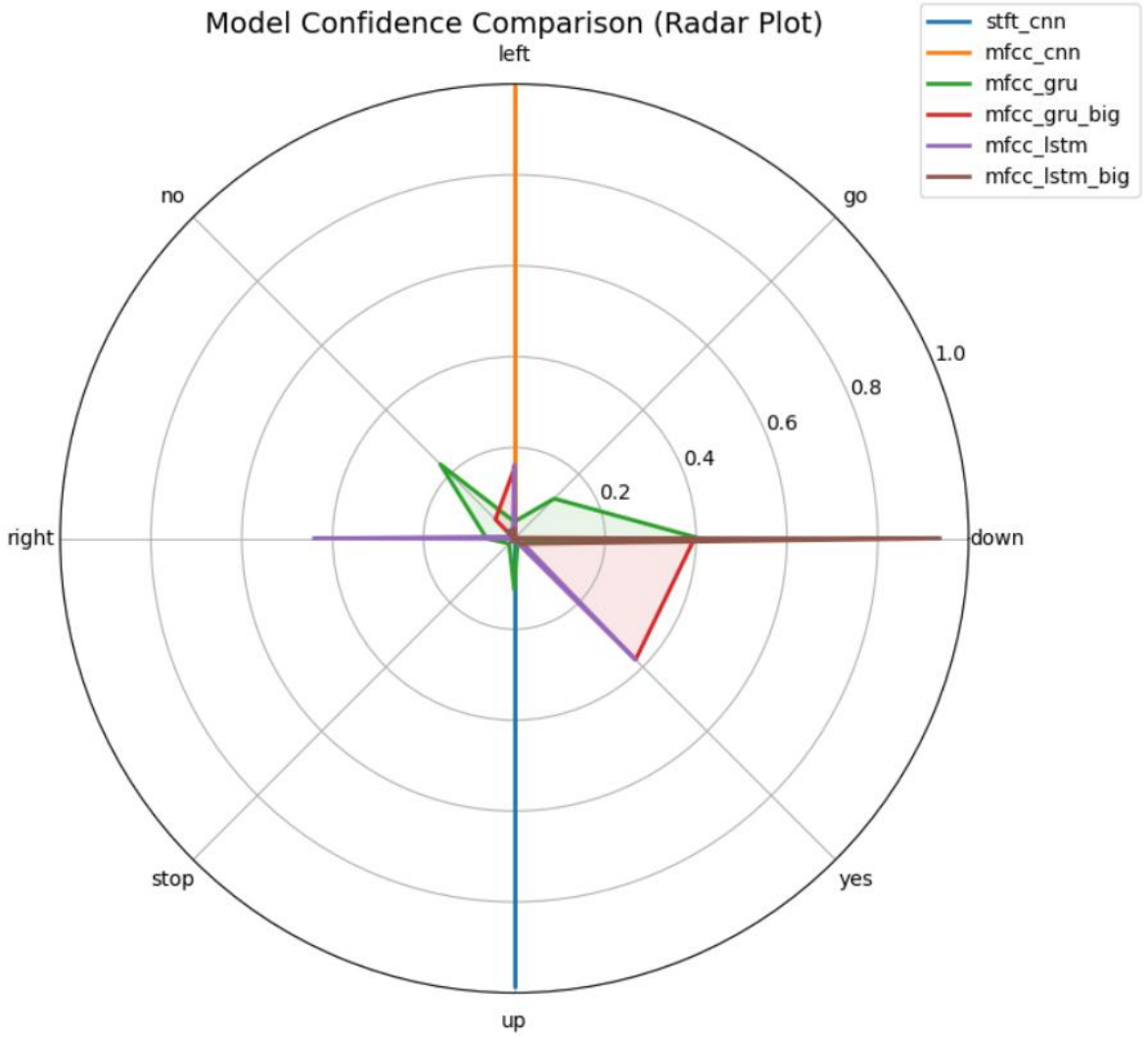


Inference Confidence

Input: left

Inference result:

```
stft_cnn: up (0.984)
mfcc_cnn: left (0.996)
mfcc_gru: down (0.408)
mfcc_gru_big: down (0.397)
mfcc_lstm: right (0.442)
mfcc_lstm_big: down (0.935)
```

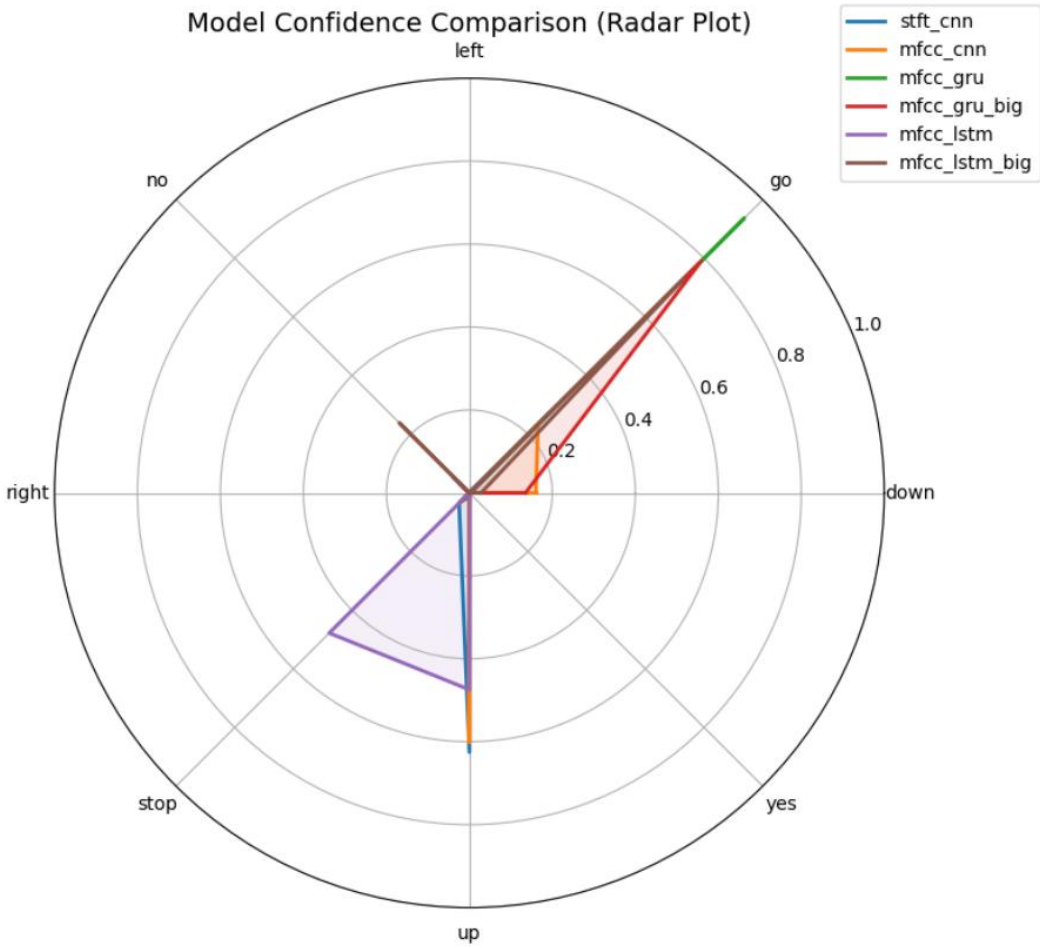


Inference Confidence

Input: go

Inference result:

```
stft_cnn: up (0.625)
mfcc_cnn: up (0.602)
mfcc_gru: go (0.937)
mfcc_gru_big: go (0.790)
mfcc_lstm: stop (0.478)
mfcc_lstm_big: go (0.730)
```



Inference Confidence

Input: down

Inference result:

```
stft_cnn: up (0.910)
mfcc_cnn: down (0.574)
mfcc_gru: down (0.860)
mfcc_gru_big: down (0.979)
mfcc_lstm: down (0.755)
mfcc_lstm_big: down (0.877)
```