

# Speech Recognition of Malayalam Numbers using Kaldi

Arunima Prasad J S  
Mtech, Dept of Electronics & Communication  
Govt. Engineering College Bartonhill, TVPM  
Kerala, India

Aparna S Thamby  
Assistant Professor, Dept of Electronics & Communication  
Govt. Engineering College Bartonhill, TVPM  
Kerala, India

**Abstract**— Speech is a straightforward and usable method of correspondence between people for trading data, however these days people aren't restricted to interfacing yet even to the various machines in our lives, most significant it is a computer. So this correspondence strategy is frequently utilized among PCs and people known as Human Computer Interface (HCI). If a system can understand human language, then it is the best method of interaction between a human and a computer. When a person speak with his own language, the machine must be able to understand it. Natural language processing (NLP) have significant role in these speech Recognition system which converts the speech from human being to text of respective language. Such systems take speech as input, recognize it and convert it into text. Malayalam is the official language of Kerala and it is spoken by 32 million people in India. Malayalam Digit Recognition have significant role in the developing technical field. The system which recognizes the Malayalam numbers spoken and convert it into text. It can be done using Deep learning technique by Kaldi toolkit. The framework utilizes Mel Frequency cepstral Coefficient (MFCC) with cepstral Mean and Variance Normalization (CMVN) and I-Vectors as feature extraction for signal handling . The framework is prepared on a test set of various qualities included dataset.

**Keywords**—: *Kaldi, MFCC, DNN*

## I. INTRODUCTION

Automatic speech recognition is a highly technical thing that makes machine turn the speech sign to the relating text or order subsequent to perceiving and understanding. ASR consolidates the extraction and confirmation of the acoustic part, the acoustic model, and the language model. The ASR can be act in any customary vernaculars, for instance, Malayalam, Tamil, Telugu, etc. In Malayalam it uses a ton of expansion [1]

Malayalam has a place with the Dravidian group of dialects and is one of the four significant dialects of this family with a rich scholarly custom. Malayalam is the mother tongue of Kerala, one of the southern state of India, and in the association region of Lakshadweep. In Malayalam language, individuals talk in various slangs. Individuals from each piece of Kerala utilize an alternate shoptalk. There are 14 regions in Kerala, and individuals live in each spot talk with various shoptalk. For every well known word individuals from every locale utilize different shoptalk to articulate it or offer something else entirely. If there should be an occurrence of Digit likewise these intricacy shows up. The framework make simple the issue. It might accommodating in numerous specialized field in which numbers can entered by talk. There

are different spoken structures in Malayalam despite the fact that the abstract lingo all through Kerala is practically uniform [3].

## II. EXISTING SYSTEM

The are many works reported on Malayalam Speech Recognition. But less work has been reported on Digit Recognition. The existing work only included the single digit values, i.e, From only Zero to Nine. For large application purpose it is not satisfactory.

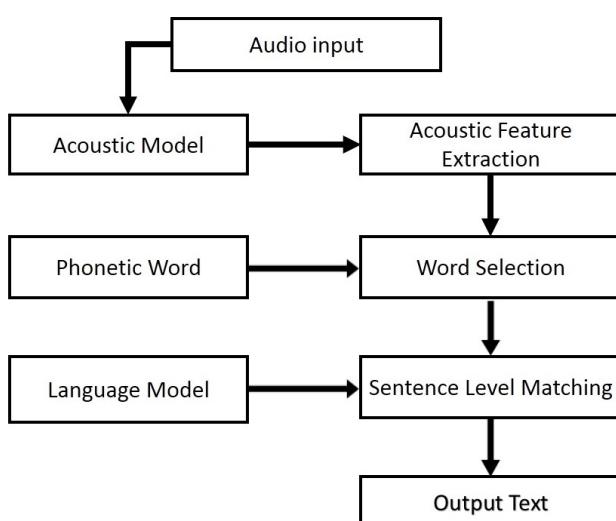
## III. MOTIVATION

Malayalam included in the twenty two scheduled languages of India. Government of India declared it as classical language in 2013. Malayalam Digit Recognition have many applications in real life. For example, the number of people who does not use ATM are less, but some people with lack of knowledge of English language and use of computer, cannot use the ATM for their needs. At that time they can use any machine by the very best communication medium, speech. They can access the machine by speaking the PIN amount, etc. With the advent of Malayalam Digit Recognition, physically challenged people will also be able to access machines like ATM's Computers etc. Therefore, designing such a helpful system like Malayalam speech recognition, especially Digit recognition will be helpful for the people and also it will be a great innovation. And also the dataset available for Malayalam Especially for Malayalam Digit is less.

## IV. METHODOLOGY

The system is an advanced Speech Recognition System for Malayalam Language which recognize Malayalam Numbers using the speech recognition tool Kaldi and Deep Neural Network (DNN). The input signal is the Dataset of Spoken Malayalam number file. i.e, Collection of recorded voices of Malayalam numbers. The system recognition is carried out by processing the speech dataset by the toolkit Kaldi which is used for testing and training purposes. The Deep Neural Network (DNN) is used for the acoustic modelling. Before the testing and training phase, the database should be feature extracted because of the need of phonetically important characteristics of speech. This can be separated by the feature extraction module which comprises MFCC, CMVN, and I-vectors.

## A. *Architecture*



### *B. Steps Included*

- Process incoming wav speech
  - From wave signal, extract acoustic features using acoustic model.
  - Linking those features to words or vocabulary or lexicon.
  - Language model or grammar defines how words can be connected to each.

## IV. KALDI

Kaldi toolkit is an open-source tool stash for speech recognition written in C++ and authorized under the Apache License v2.0. The objective of Kaldi is to have a versatile code that is direct, alter and expand. The instruments total on the for the most part used Unix-like structures and on Microsoft Windows. Kaldi is conveyed under the Apache License v2.0, which is significantly nonrestrictive, making it sensible for a wide neighborhood clients [5].

## A. *Features*

- Reconciliation with Finite State Transducers
  - Broad linear algebra support
  - BLAS and LAPACK schedule
  - Extensible plan: scores.
  - Open permit
  - Complete recipes
  - Careful testing:

## *B. Parts of Kaldi*

- 1) Preprocessing and Feature Extraction
  - 2) Modelling
  - 3) Training

## V. DATA COLLECTION

#### A. *Speech Data Collection*

For the preparation of spoken digit data, the speech is recorded in both noise and noiseless environment.

The framework is intended to perceive Malayalam digits, therefore the length or size of the vocabulary is eleven (counting quiet). For noisy and noiseless environment, a top quality receiver with headset can be utilized which includes mouthpiece with 70Hz to 16000 Hz of recurrence range. The recording should be possible with 16 kHz inspecting recurrence quantized by 16 bit. To catch all the acoustic property of the words, the training dataset is intended. A Transcription document is made for every expression of the speaker and a language word reference is made for each word in the string. These are stored in discrete records. For noisy environment, the recording manner is same as that of the recording in the noiseless environment. The recording device used is show in the next section. It is better and advantageous.

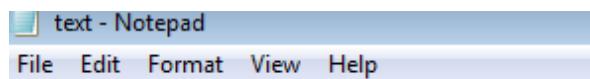
## *B. Recording*

Recording is done in both noisy and noiseless environment. In both cases the recording device Roland R-07 is used because it can capture both noisy and noiseless audios. The numbers from zero to 9999 are recorded with files groups. The specifications used for recording are

- Frequency = 48KHz
  - Sampling Rate = 16 bit
  - Recording of each audio file is saved as .wav format
  - The editing includes noise reduction etc. are done by using the Software Audacity.
  - Recording is done with 10 different persons with 5 male and 5 female with different age groups and slangs.

### C. *Text Data Collection*

The collection of text data is done by creating a text file with the text of numbers spoken by the speakers. Fig. Shows an example



என் நன்கு முன் நால் அன்றை  
ஏதும் ஏதும் என்பது புகை  
பறத் தனிகொள் புறநெங்கு பதிமுள் பதிகொல்  
பதிகொல்க் கதிகொர் பதிகொஷ் பதிகொக்கு பதிதாக்கபத்  
ஹங்கபத் முழுத் நால்பத் அங்கபத் அங்கபத்  
ஏதும்பத் ஏதும்பத் தொழைந் டூர்  
ஹங்கபதி முபுறதி நால்பதி அங்கபதி அங்கபதி  
ஏதும்பதி ஏதும்பதி தொழைந்தி டூடி  
ஹங்காந் முகாந் நாகாந் அங்காந் அங்காந்  
ஏதும்காந் ஏதும்காந் தொகுதாயிரங் குடியிரங் |

Each line in the transcript file corresponds to each text file with respect to the recorded audio file.

## VI. FEATURE EXTRACTION

### A. Mel Frequency Cepstral coefficient (MFCC)

MFCC is a popular technique for extracting features. To catch the phonetically significant attributes of speech, Mel Frequency scale is used to express the signal. This scale has below 1000Hz linearly frequency spacing and a above 1000Hz logarithmic spacing. MFCCs are less involves to the states of being of the speakers' vocal rope, contrasted with the speech wave structures.

### B. Cepstral Mean and Variance Normalization (CMVN)

Cepstral mean and fluctuation standardization (CMVN) is a computationally proficient standardization strategy for strong speech acknowledgment. CMVN minimizes noise distortion for robust feature extraction by linearly transforming the cepstral coefficients to have the same segmental statistics

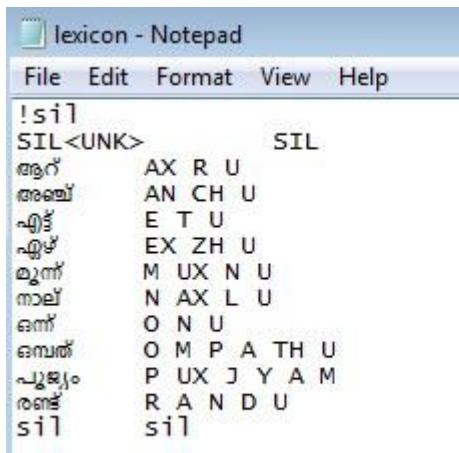
### C. I-Vectors

I-Vectors are utilized for addressing the style of every sound expression or speaker. I-Vectors are utilized for better comprehension of the changes inside the kaldi.

## VII. EXPERIMENTAL ANALYSIS

### A. Creation of Lexicon file

To train the Malayalam Database, a lexicon file is needed. The lexicon file is the phonetic transcription of each number



### B. Training of data

The training of data is done by eleven different speakers with both noise and noiseless environment. The Malayalam numbers are from zero to 99999. The overall WER is 35 %

### C. Word Error Rate per Speakers

The fig shows the WER per each speaker. The WER can be varies according to the speaker with their sound, voice, slang etc. and also depends on the environment i.e, noisy or noiseless.

SPK SPEAKER	id	#SENT	#WORD	Corr	Sub	Ins	Del	Err	S.Err
ML001	raw	2	14	21.43	0.00	0.00	78.57	78.57	100.00
ML001	sys	2	14	0	0	0	14	14	2
ML002	raw	2	14	0.00	0.00	0.00	100.00	100.00	100.00
ML003	raw	2	14	1	2	0	11	13	2
ML003	sys	2	14	7.14	14.29	0.00	78.57	92.86	100.00
ML004	raw	2	6	1	0	0	5	5	2
ML004	sys	2	6	16.67	0.00	0.00	83.33	83.33	100.00
ML005	raw	5	25	20	2	2	3	7	5
ML005	sys	5	25	80.00	8.00	8.00	12.00	28.00	100.00
ML006	raw	3	15	11	3	0	1	4	3
ML006	sys	3	15	73.33	20.00	0.00	6.67	26.67	100.00
ML007	raw	3	15	7	6	1	2	9	3
ML007	sys	3	15	46.67	40.00	6.67	13.33	60.00	100.00
ML009	raw	3	13	10	0	0	3	3	3
ML009	sys	3	13	76.92	0.00	0.00	23.08	23.08	100.00
ML010	raw	3	14	11	0	2	3	5	3
ML010	sys	3	14	78.57	0.00	14.29	21.43	35.71	100.00
ML011	raw	3	15	11	2	0	2	4	3
ML011	sys	3	15	73.33	13.33	0.00	13.33	26.67	100.00
SUM	raw	28	145	75	15	5	55	75	28
SUM	sys	28	145	51.72	10.34	3.45	37.93	51.72	100.00

FIG WER PER SPEAKER

### D. Testing of some examples

The fig shows the testing of some example datasets. It shows 75% accuracy

```
input1 ഒരുപദി അപ്പോരുന്നും അപ്പോരുന്നും നാലുപദി അപ്പോരുന്നും
input2 ഒന്നും അപ്പോരുന്നും അപ്പോരുന്നും
input3 ഒന്നും അപ്പോരുന്നും
input4 ഒന്നും അപ്പോരുന്നും
input5 ഒന്നും അപ്പോരുന്നും
input6 ഒന്നും അപ്പോരുന്നും
input7 ഒന്നും അപ്പോരുന്നും
input8 ഒന്നും അപ്പോരുന്നും
input9 ഒന്നും അപ്പോരുന്നും
input10 ഒന്നും അപ്പോരുന്നും
input11 ഒന്നും അപ്പോരുന്നും
input12 ഒന്നും അപ്പോരുന്നും
input13 ഒന്നും അപ്പോരുന്നും
input14 ഒന്നും അപ്പോരുന്നും
input15 ഒന്നും അപ്പോരുന്നും
input16 ഒന്നും അപ്പോരുന്നും
input17 ഒന്നും അപ്പോരുന്നും
input18 ഒന്നും അപ്പോരുന്നും
input19 ഒന്നും അപ്പോരുന്നും
input20 ഒന്നും അപ്പോരുന്നും
input21 ഒന്നും അപ്പോരുന്നും
input22 ഒന്നും അപ്പോരുന്നും
input23 ഒന്നും അപ്പോരുന്നും
input24 ഒന്നും അപ്പോരുന്നും
input25 ഒന്നും അപ്പോരുന്നും
input26 ഒന്നും അപ്പോരുന്നും
input27 ഒന്നും അപ്പോരുന്നും
input28 ഒന്നും അപ്പോരുന്നും
input29 ഒന്നും അപ്പോരുന്നും
input30 ഒന്നും അപ്പോരുന്നും
input31 ഒന്നും അപ്പോരുന്നും
input32 ഒന്നും അപ്പോരുന്നും
input33 ഒന്നും അപ്പോരുന്നും
input34 ഒന്നും അപ്പോരുന്നും
input35 ഒന്നും അപ്പോരുന്നും
input36 ഒന്നും അപ്പോരുന്നും
input37 ഒന്നും അപ്പോരുന്നും
input38 ഒന്നും അപ്പോരുന്നും
input39 ഒന്നും അപ്പോരുന്നും
input40 ഒന്നും അപ്പോരുന്നും
input41 ഒന്നും അപ്പോരുന്നും
input42 ഒന്നും അപ്പോരുന്നും
input43 ഒന്നും അപ്പോരുന്നും
input44 ഒന്നും അപ്പോരുന്നും
input45 ഒന്നും അപ്പോരുന്നും
input46 ഒന്നും അപ്പോരുന്നും
input47 ഒന്നും അപ്പോരുന്നും
input48 ഒന്നും അപ്പോരുന്നും
input49 ഒന്നും അപ്പോരുന്നും
input50 ഒന്നും അപ്പോരുന്നും
input51 ഒന്നും അപ്പോരുന്നും
input52 ഒന്നും അപ്പോരുന്നും
input53 ഒന്നും അപ്പോരുന്നും
input54 ഒന്നും അപ്പോരുന്നും
input55 ഒന്നും അപ്പോരുന്നും
input56 ഒന്നും അപ്പോരുന്നും
input57 ഒന്നും അപ്പോരുന്നും
input58 ഒന്നും അപ്പോരുന്നും
input59 ഒന്നും അപ്പോരുന്നും
input60 ഒന്നും അപ്പോരുന്നും
input61 ഒന്നും അപ്പോരുന്നും
input62 ഒന്നും അപ്പോരുന്നും
input63 ഒന്നും അപ്പോരുന്നും
input64 ഒന്നും അപ്പോരുന്നും
input65 ഒന്നും അപ്പോരുന്നും
input66 ഒന്നും അപ്പോരുന്നും
input67 ഒന്നും അപ്പോരുന്നും
input68 ഒന്നും അപ്പോരുന്നും
input69 ഒന്നും അപ്പോരുന്നും
input70 ഒന്നും അപ്പോരുന്നും
input71 ഒന്നും അപ്പോരുന്നും
input72 ഒന്നും അപ്പോരുന്നും
input73 ഒന്നും അപ്പോരുന്നും
input74 ഒന്നും അപ്പോരുന്നും
input75 ഒന്നും അപ്പോരുന്നും
input76 ഒന്നും അപ്പോരുന്നും
input77 ഒന്നും അപ്പോരുന്നും
input78 ഒന്നും അപ്പോരുന്നും
input79 ഒന്നും അപ്പോരുന്നും
input80 ഒന്നും അപ്പോരുന്നും
input81 ഒന്നും അപ്പോരുന്നും
input82 ഒന്നും അപ്പോരുന്നും
input83 ഒന്നും അപ്പോരുന്നും
input84 ഒന്നും അപ്പോരുന്നും
input85 ഒന്നും അപ്പോരുന്നും
input86 ഒന്നും അപ്പോരുന്നും
input87 ഒന്നും അപ്പോരുന്നും
input88 ഒന്നും അപ്പോരുന്നും
input89 ഒന്നും അപ്പോരുന്നും
input90 ഒന്നും അപ്പോരുന്നും
input91 ഒന്നും അപ്പോരുന്നും
input92 ഒന്നും അപ്പോരുന്നും
input93 ഒന്നും അപ്പോരുന്നും
input94 ഒന്നും അപ്പോരുന്നും
input95 ഒന്നും അപ്പോരുന്നും
input96 ഒന്നും അപ്പോരുന്നും
input97 ഒന്നും അപ്പോരുന്നും
input98 ഒന്നും അപ്പോരുന്നും
input99 ഒന്നും അപ്പോരുന്നും
input100 ഒന്നും അപ്പോരുന്നും
input101 ഒന്നും അപ്പോരുന്നും
input102 ഒന്നും അപ്പോരുന്നും
input103 ഒന്നും അപ്പോരുന്നും
input104 ഒന്നും അപ്പോരുന്നും
input105 ഒന്നും അപ്പോരുന്നും
input106 ഒന്നും അപ്പോരുന്നും
input107 ഒന്നും അപ്പോരുന്നും
input108 ഒന്നും അപ്പോരുന്നും
input109 ഒന്നും അപ്പോരുന്നും
input110 ഒന്നും അപ്പോരുന്നും
input111 ഒന്നും അപ്പോരുന്നും
input112 ഒന്നും അപ്പോരുന്നും
input113 ഒന്നും അപ്പോരുന്നും
input114 ഒന്നും അപ്പോരുന്നും
input115 ഒന്നും അപ്പോരുന്നും
input116 ഒന്നും അപ്പോരുന്നും
input117 ഒന്നും അപ്പോരുന്നും
input118 ഒന്നും അപ്പോരുന്നും
input119 ഒന്നും അപ്പോരുന്നും
input120 ഒന്നും അപ്പോരുന്നും
input121 ഒന്നും അപ്പോരുന്നും
input122 ഒന്നും അപ്പോരുന്നും
input123 ഒന്നും അപ്പോരുന്നും
input124 ഒന്നും അപ്പോരുന്നും
input125 ഒന്നും അപ്പോരുന്നും
input126 ഒന്നും അപ്പോരുന്നും
input127 ഒന്നും അപ്പോരുന്നും
input128 ഒന്നും അപ്പോരുന്നും
input129 ഒന്നും അപ്പോരുന്നും
input130 ഒന്നും അപ്പോരുന്നും
input131 ഒന്നും അപ്പോരുന്നും
input132 ഒന്നും അപ്പോരുന്നും
input133 ഒന്നും അപ്പോരുന്നും
input134 ഒന്നും അപ്പോരുന്നും
input135 ഒന്നും അപ്പോരുന്നും
input136 ഒന്നും അപ്പോരുന്നും
input137 ഒന്നും അപ്പോരുന്നും
input138 ഒന്നും അപ്പോരുന്നും
input139 ഒന്നും അപ്പോരുന്നും
input140 ഒന്നും അപ്പോരുന്നും
input141 ഒന്നും അപ്പോരുന്നും
input142 ഒന്നും അപ്പോരുന്നും
input143 ഒന്നും അപ്പോരുന്നും
input144 ഒന്നും അപ്പോരുന്നും
input145 ഒന്നും അപ്പോരുന്നും
input146 ഒന്നും അപ്പോരുന്നും
input147 ഒന്നും അപ്പോരുന്നും
input148 ഒന്നും അപ്പോരുന്നും
input149 ഒന്നും അപ്പോരുന്നും
input150 ഒന്നും അപ്പോരുന്നും
input151 ഒന്നും അപ്പോരുന്നും
input152 ഒന്നും അപ്പോരുന്നും
input153 ഒന്നും അപ്പോരുന്നും
input154 ഒന്നും അപ്പോരുന്നും
input155 ഒന്നും അപ്പോരുന്നും
input156 ഒന്നും അപ്പോരുന്നും
input157 ഒന്നും അപ്പോരുന്നും
input158 ഒന്നും അപ്പോരുന്നും
input159 ഒന്നും അപ്പോരുന്നും
input160 ഒന്നും അപ്പോരുന്നും
input161 ഒന്നും അപ്പോരുന്നും
input162 ഒന്നും അപ്പോരുന്നും
input163 ഒന്നും അപ്പോരുന്നും
input164 ഒന്നും അപ്പോരുന്നും
input165 ഒന്നും അപ്പോരുന്നും
input166 ഒന്നും അപ്പോരുന്നും
input167 ഒന്നും അപ്പോരുന്നും
input168 ഒന്നും അപ്പോരുന്നും
input169 ഒന്നും അപ്പോരുന്നും
input170 ഒന്നും അപ്പോരുന്നും
input171 ഒന്നും അപ്പോരുന്നും
input172 ഒന്നും അപ്പോരുന്നും
input173 ഒന്നും അപ്പോരുന്നും
input174 ഒന്നും അപ്പോരുന്നും
input175 ഒന്നും അപ്പോരുന്നും
input176 ഒന്നും അപ്പോരുന്നും
input177 ഒന്നും അപ്പോരുന്നും
input178 ഒന്നും അപ്പോരുന്നും
input179 ഒന്നും അപ്പോരുന്നും
input180 ഒന്നും അപ്പോരുന്നും
input181 ഒന്നും അപ്പോരുന്നും
input182 ഒന്നും അപ്പോരുന്നും
input183 ഒന്നും അപ്പോരുന്നും
input184 ഒന്നും അപ്പോരുന്നും
input185 ഒന്നും അപ്പോരുന്നും
input186 ഒന്നും അപ്പോരുന്നും
input187 ഒന്നും അപ്പോരുന്നും
input188 ഒന്നും അപ്പോരുന്നും
input189 ഒന്നും അപ്പോരുന്നും
input190 ഒന്നും അപ്പോരുന്നും
input191 ഒന്നും അപ്പോരുന്നും
input192 ഒന്നും അപ്പോരുന്നും
input193 ഒന്നും അപ്പോരുന്നും
input194 ഒന്നും അപ്പോരുന്നും
input195 ഒന്നും അപ്പോരുന്നും
input196 ഒന്നും അപ്പോരുന്നും
input197 ഒന്നും അപ്പോരുന്നും
input198 ഒന്നും അപ്പോരുന്നും
input199 ഒന്നും അപ്പോരുന്നും
input200 ഒന്നും അപ്പോരുന്നും
input201 ഒന്നും അപ്പോരുന്നും
input202 ഒന്നും അപ്പോരുന്നും
input203 ഒന്നും അപ്പോരുന്നും
input204 ഒന്നും അപ്പോരുന്നും
input205 ഒന്നും അപ്പോരുന്നും
input206 ഒന്നും അപ്പോരുന്നും
input207 ഒന്നും അപ്പോരുന്നും
input208 ഒന്നും അപ്പോരുന്നും
input209 ഒന്നും അപ്പോരുന്നും
input210 ഒന്നും അപ്പോരുന്നും
input211 ഒന്നും അപ്പോരുന്നും
input212 ഒന്നും അപ്പോരുന്നും
input213 ഒന്നും അപ്പോരുന്നും
input214 ഒന്നും അപ്പോരുന്നും
input215 ഒന്നും അപ്പോരുന്നും
input216 ഒന്നും അപ്പോരുന്നും
input217 ഒന്നും അപ്പോരുന്നും
input218 ഒന്നും അപ്പോരുന്നും
input219 ഒന്നും അപ്പോരുന്നും
input220 ഒന്നും അപ്പോരുന്നും
input221 ഒന്നും അപ്പോരുന്നും
input222 ഒന്നും അപ്പോരുന്നും
input223 ഒന്നും അപ്പോരുന്നും
input224 ഒന്നും അപ്പോരുന്നും
input225 ഒന്നും അപ്പോരുന്നും
input226 ഒന്നും അപ്പോരുന്നും
input227 ഒന്നും അപ്പോരുന്നും
input228 ഒന്നും അപ്പോരുന്നും
input229 ഒന്നും അപ്പോരുന്നും
input230 ഒന്നും അപ്പോരുന്നും
input231 ഒന്നും അപ്പോരുന്നും
input232 ഒന്നും അപ്പോരുന്നും
input233 ഒന്നും അപ്പോരുന്നും
input234 ഒന്നും അപ്പോരുന്നും
input235 ഒന്നും അപ്പോരുന്നും
input236 ഒന്നും അപ്പ
```

- 
- [6] Arun HP, S. R. A. S. A., Jithin Kunjumon (2021). Malayalam speech to text conversion using deep learning. IOSR Journal of Engineering (IOSRJEN), 11, 24–30
  - [7] B, B. L., G. Anu, K. R. Sreelakshmi, and L. Mary, Continuous speech recognition system for malayalam language using kald. In 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR).2018.
  - [8] Moncy, A. M., A. M., H. Jasmin, and R. Rajan, Automatic speech recognition in malayalam using dnn-based acoustic modelling. In 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS). 2020.
  - [9] S., R., A. Joseph, and A. B. K.K., Isolated digit recognition for malayalam- an application perspective. In 2013 International Conference on Control Communication and Computing (ICCC). 2013.