
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Developing a Speech Recognition System for the Urdu Language

Members:

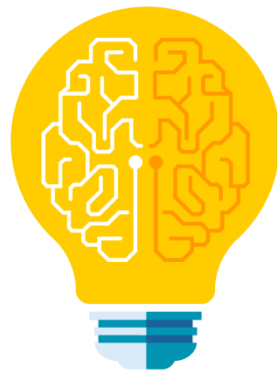
Muhammad Ibtesam Arshad

FA17-BSE-061

Hamza Safdar

FA17-BSE-013





Speech Recognition System for the Urdu Language

OUTLINE

Data Collection and Model Training

Build Flask API

Deployment of Flask API

Development of Mobile Application

Data Collection

DATA:

**Center for Language Engineering Speech
Corpus**

Name:

Phonetically Rich Urdu Speech

Corpus

Total Instances:

708 Speeches

Split Dataset

DATASET

The Dataset used for this Project comprises of

- **708 Instances**

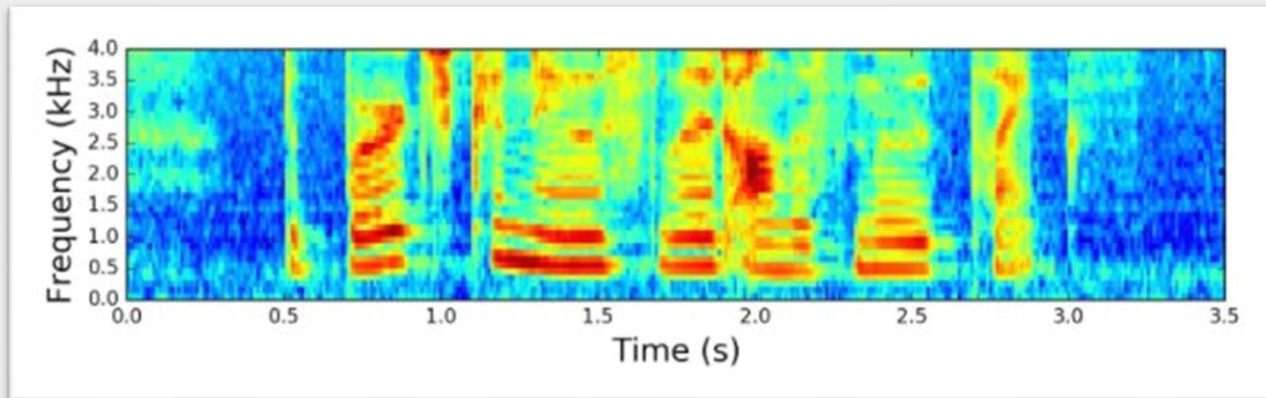
Sample Data Characteristics

- Total Instances in Sample Data = 708

Training = 658 and Testing = 50

Data Preparation

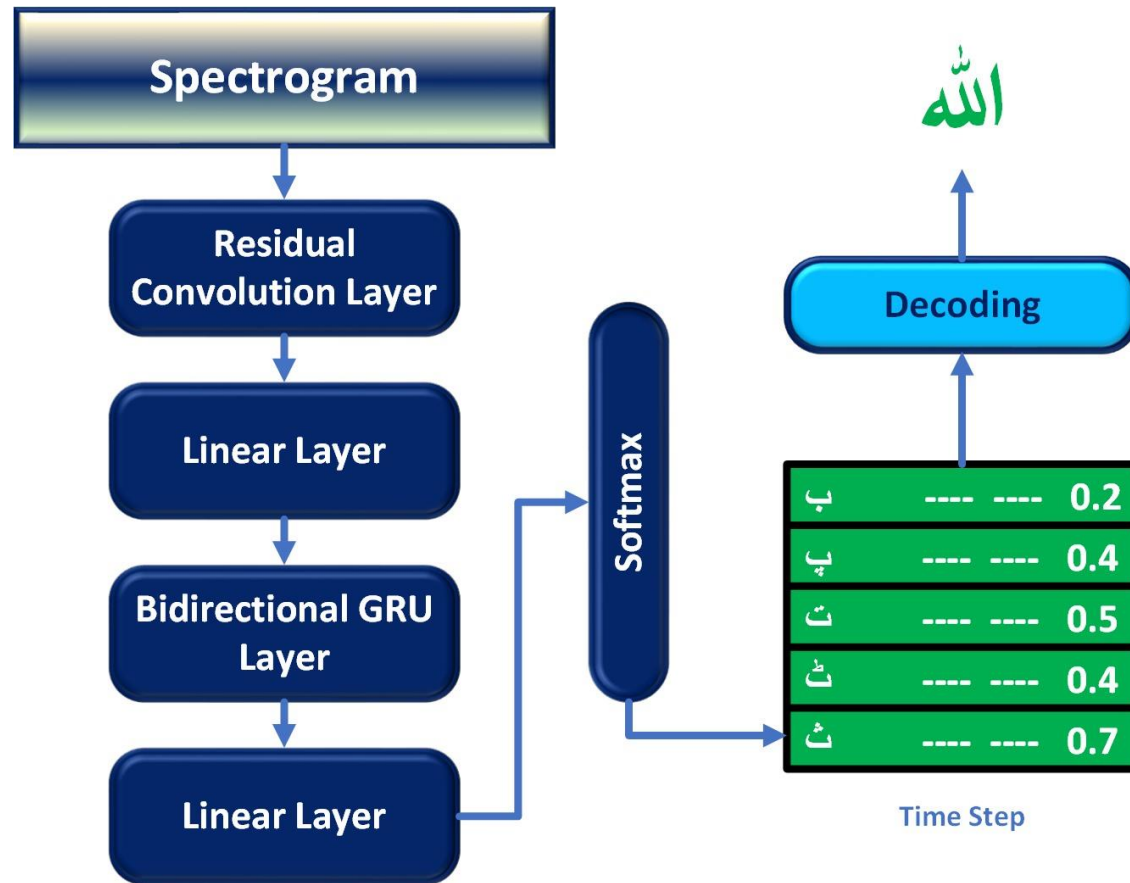
We'll take raw audio waves and transform them into Mel Spectrograms.



Model Architecture

Our model will be similar to the Deep Speech 2 architecture. The model will have two main neural network modules - N layers of Residual Convolutional Neural Networks (ResCNN) to learn the relevant audio features, and a set of Bidirectional Recurrent Neural Networks (BiRNN) to leverage the learned ResCNN audio features. The model is topped off with a fully connected layer used to classify characters per time step.

Model Architecture



Evaluating Speech Model

- We used a **Greedy Decoding** method to process our model's output into characters that can be combined to create the transcript.
- A **Greedy Decoder** takes in the model output, which is a softmax probability matrix of characters, and for each time step (spectrogram frame), it chooses the label with the highest probability. If the label is a blank label, we remove it from the final transcript.
- Then we calculated:
 - **Word Error** and **Character Error** using Levenshtein Distance.

Deploying Model through FLASK

FLASK Framework

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.

Routing



1

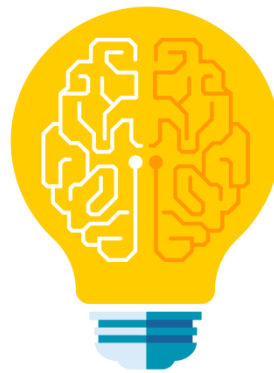
```
# Route to handle HOME
@app.route('/')
def home():
```

2

```
# Route to handle PREDICTED RESULT
@app.route('/predict')
def predict():
```

3

```
# Route to Handle API Request
@app.route('/predict-api')
def predict_api():
```



Interfaces

Web Application

and

Mobile Application

Landing Page – Web App



Speech Recognition System for the Urdu Language

Asslam-o-Alaikum !

Automatic Speech Recognition Systems (ASR) are used to convert the acoustic signals which are caught through the microphones into the sequence of words. The Automatic Speech Recognition Systems (ASR) empower the machines to react correctly, reliably, and effectively to human speech or voice and offer helpful and important services to the users. The main Purpose of this project is Building an end-to-end Speech Recognition System in PyTorch on CLE publically available Phonetically Rich Urdu Speech Corpus.

[Read Instructions](#)



Select a file to upload that you desire to
convert the text to speech

Copyrights © 2020 **Muhammad Ibtesam Arshad** | All Rights Reserved

Instructions – Web App



Speech Recognition for the Urdu


Asslam-o-Alaikum !

Automatic Speech Recognition Systems (ASR) convert acoustic signals which are caught through the microphones into the sequence of words. The Automatic Speech Recognition Systems (ASR) empower the machines to react correctly, reliably, and effectively to human speech or voice and offer helpful and important services to the users. The main Purpose of this project is Building an end-to-end Speech Recognition System in PyTorch on CLE publicly available Phonetically Rich Urdu Speech Corpus.

[Read Instructions](#)

Steps to Follow

- Step 1 :** Click on **Upload button** to open file chooser
- Step 2 :** Select the **Audio File (.wav)** from file chooser
- Step 3 :** Your **selected file** will be passed to our model as well as to google API
- Step 4 :** The **Predicted Text** will be displayed back to you
- Step 5 :** You can **compare** the **predicted text** of both models



Select a file to upload that you desire to convert the text to speech

Copyrights © 2020 **Muhammad Ibtesam Arshad** | All Rights Reserved

Predictions – Web App



Speech Recognition for the Urdu

Assalam-o-Alaikum !

Automatic Speech Recognition System processes acoustic signals which are caught by a microphone and converts them into a sequence of words. The Automatic Speech Recognition System empowers the machines to react to speech or voice and offer helpful answers.

[Read Instructions](#)

Prediction of Model

Selected Audio

0:07 / 0:07

Our Model Prediction

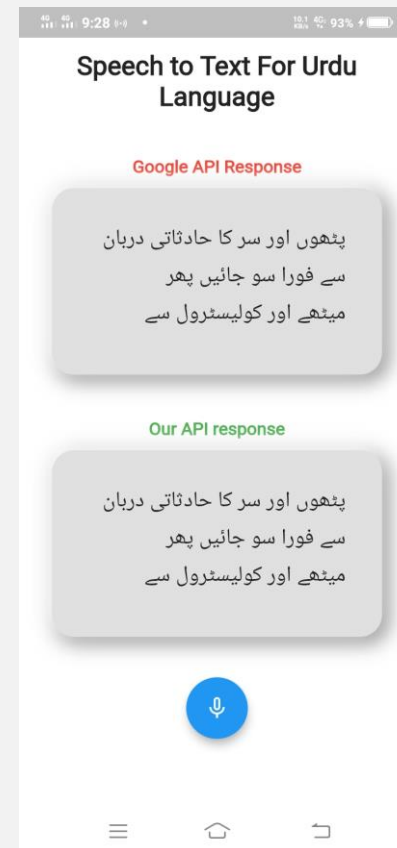
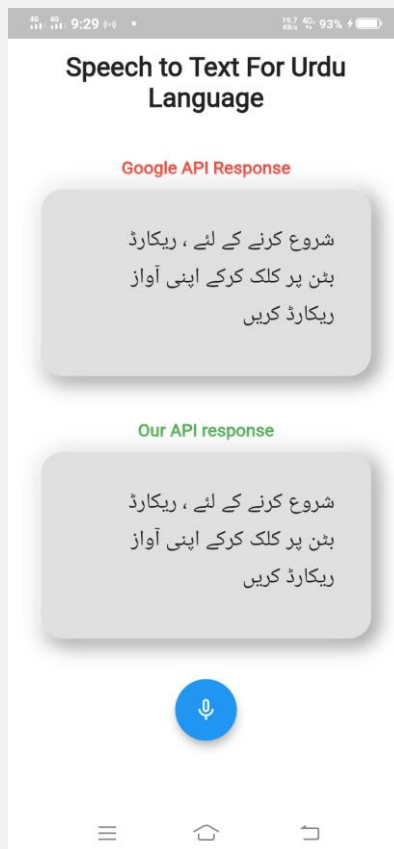
پٹھوں اور سرکا کا دھاتی درد بام سے فوراً سلجھائیں پھر میٹھے اور
کولیٹرول سے بچیں

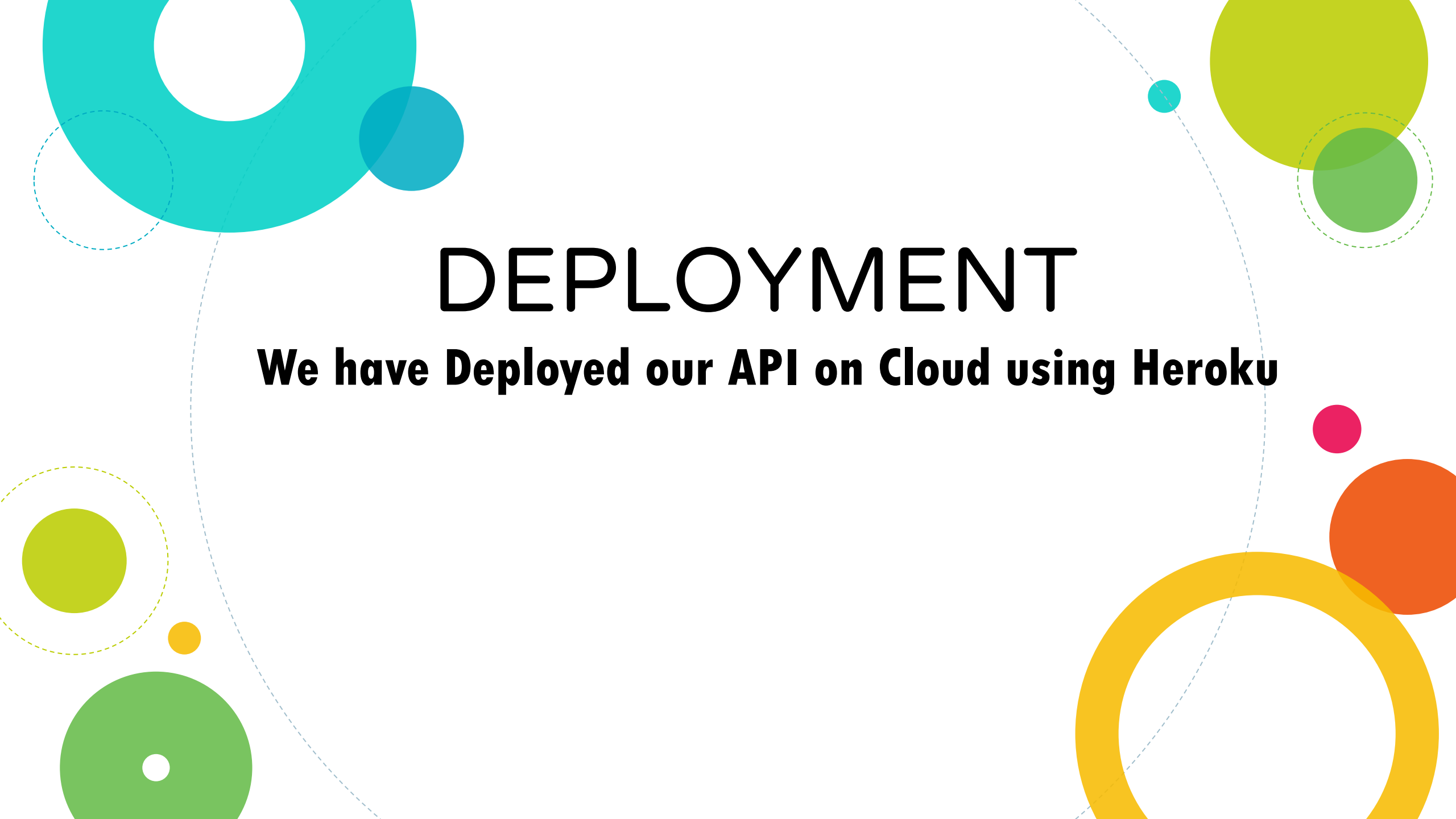
Google API Prediction

پٹھوں اور سرکا کا دھاتی درد بام سے فوراً سو جائیں پھر میٹھے اور
کولیٹرول سے بچیں

Click to upload that you desire to convert the text to speech

Predictions – Mobile App



The background features a collection of colorful circles in shades of teal, lime green, orange, and pink. Some circles are solid, while others are dashed outlines. A thin, light blue dashed line curves across the slide, passing through several of the circles.

DEPLOYMENT

We have Deployed our API on Cloud using Heroku

API Response



```
{
  Gender: "Male",
  - Predicted Text: {
    Our Prediction: "پٹھوں اور سرکا حات کدارد بام سے فوراً سلجھائیں پھر میٹھے اور کولیسٹرول سے بچیں",
    Roman: "Pathar aur sar ka Hadsa t Dard Bam se foreign suljhaen FIR Mithe aur cholesterol se bache",
    Urdu: "پٹھوں اور سر کا حادثاتی درد بام سے فوراً سو جائیں پھر میٹھے اور کولیسٹرول سے بچیں"
  }
}
```

[+ - View source](#)

An abstract graphic design featuring a white background with a large, faint, light blue dashed circle. Various colorful circles are scattered around the perimeter of this dashed circle. In the top-left, there is a large teal circle with a white center, a smaller solid teal circle, and a dashed teal circle. In the top-right, there is a large lime green circle, a smaller solid green circle, and a dashed green circle. In the bottom-left, there is a large green circle with a white center, a smaller solid green circle, and a dashed green circle. In the bottom-right, there is a large yellow circle, a smaller solid orange circle, a small solid pink circle, and a small solid orange circle. The text "FUTURE WORK" is centered in the middle of the image in a bold, black, sans-serif font.

FUTURE WORK

Future Work

TASK 1

Speech Recognition
System for the
English Language

TASK 2

Speech Recognition
System for the
Urdu Language

جَزَاكَ اللَّهُ خَيْرًا
