Medical Transcription for Electronic Healthcare Record (EHR) in NLP

Project by

Aditya Wadnerker [014428557]
Jacky Chow [014545284]
Jerry Xianghui Huang [014516905]
Jumana Nadir [014502917]

GitHub Repository-

https://github.com/zjzsu2000/CMPE256_project

Medical Transcription for Electronic Healthcare Record (EHR) in NLP

Abstract

Speech recognition has boomed today because of advanced NLP resources and interdisciplinary technology. Different API resources makes it possible for easy and accurate Speech to Text conversion. However, Medical Transcription remains challenging and expensive because of the difficult lexical and Medical Terminology. There are a lot of EHR (Electronic Healthcare Record) softwares available today which offer machine-led speech recognition, but accuracy is a prevalent question. Another commonly used technique is the human-led transcription, which has higher accuracy but is a strenuous job. In this project, we have attempted to develop a Speech Recognition model for Medical Dictation for making an EHR.

Introduction

In recent years, the technology of electronic healthcare record is more and more widely applied in the field of medical treatment. With the continuous development of the data science, conversational assistants are also developing from the original text recognition to Artificial intelligent assistants. Artificial intelligent assistants as an assistant with the diagnosis and treatment process, the main advantage is high velocity, high accurate and the low error rate. AI assistants help doctors deal with a large amount of tedious and cumbersome repetitive works. Also reduce the burden of the doctors. With the improvement of technology and innovation, we strongly believe that, conversation as the medium of human-computer interaction is more likely to become the next generation of meaningful interaction way of healthcare industry.

The first generation of conversation technology is used to solve various problems exist in the process of medical treatment. For example, a company in the US is the use of conversation recognition technology has invented an in-ear equipment, with this device, the surgeon can directly communicate with the electronic medical record of the patient. And get the more accurate information about the patient compared to asking patients themselves. By using this equipment, it saves a lot of time for the doctor and effectively solved the doctor shortage for that time.

Electronic Healthcare record (EHR) is one of the most important and successful cases in medical conversation technology. Many companies are doing the project about EMR in today's world. Amazon developed Alexa platform, it is a conversational assistant that can make communication between doctors and patients. And it can also send the alarm when the patients in some emergency situation. Alexa also have a strong healthcare management system that can manage lots of data by the information recorded by the conversation. In some backward countries, the hospital and the doctor are not enough. Sometimes, the disease can make a huge damage to these countries. By using the EHR it can help doctors to reduce lots of time, so the doctor can treat more patients. EHR also have a strong market in China, Cause of the huge population, in some

large cities the people need to wait around two to three months for seeing the doctor, by using the EHR it can help doctors to treat more patients and reduce the time for waiting in the line.

Meanwhile, the medical conversation technology can recognize the voice in some medical scenarios like cough, Parkinson's and some autism patients, and make the voice into the easy-understanding information. Social isolation caused by communication barriers is a very serious social problem in today's world. The disability of communication can destroy people's lives, and even cause other health problems, such as depression or early death. Medical conversation technology can change the problems and change the lives of many people.

Motivation

Human-led transcriptions are extremely expensive, time consuming and are subject to inaccuracy. More and more companies are inclining towards making successful speech recognition systems in the healthcare industry. Successful voice assistants and transcription systems can help doctors and patients in many ways. Amazon has recently introduced a Medical Transcription API which can be helpful in many ways. In our research paper we have discussed various chatbots and virtual assistants and their performances. And we understand the need to create more accurate transcription tools for medical conversations.

March 2018 | Healthcare IT News | www.HealthcareITNews.com

COVER STORY 7

SPECIAL REPORT:

AI voice assistants have officially arrived in healthcare

Amazon has an early lead, but lessons from cutting-edge providers can also be applied to Apple's Siri, Google Home and Assistant and Microsoft Cortana.

By Bill Siwicki, Managing Editor

As 2017 WOUND DOWN, Amazon's Alexa voice assistant rose up to the top of Apple's App Store, while Amazon's Echo Dot was its own bestseller.

But some of the millions of those devices sold have already found their way into hospitals such as Beth Israel Deaconess Medical Center in Boston, Northwell Health in New York, the Commonwealth Care Alliance in Boston, and Libertana Home Health in Los Angeles.

"When we went from laptops to smartphones as our primary means of doing computing, that was a major paradigm shift," said John Halamka, MD, CIO at Beth Israel. "Ambient listening tools probably will replace mobile devices."

Amazon is hardly alone in the market. Apple Siri, Google Home and Assistant, and Microsoft Cortana are also available. Among hospitals undertaking pilot projects, Amazon reigns supreme — for now. At least one hospital is already girding to implement Google Home as well.

EARLY WORK AT BIDMC

Halamka's team is conducting extensive early work with Amazon's Alexa voice assistant.

Halamka's team has built a variety of skills for the Alexa voice assistant. A skill is the computing manifestation of a task that a voice assistant performs in the real world. Halamka said the Alexa APIs used to build skills are straightforward and easy to use. Beth Israel Deaconess System now has a variety of skills used in the voice assistants in patient rooms.

"If you are an inpatient, what are the things typically users identify the wait times at emergency rooms and urgent care centers that are near a given ZIP code.

"You can have Alexa ask Northwell for the shortest wait time near your ZIP code or check the time at a specific location," said Emily Kagan Trenchard, vice president of digital and innovation strategy at Northwell Health. "The skill will give you the wait time as well as the address for that location and direct users to an alternative if the shortest wait time is elsewhere."

It's also smart enough to handle disambiguation around places with similar names. For example, Northwell has three hospitals known as Long Island Jewish—Long Island Jewish Valley Stream, Long Island Jewish Forest Hills and Long Island Jewish Medical Center.

"Imagine a scenario where you are in the kitchen and you cut your finger while cooking," Trenchard said. "In this instance, you know you need to get medical attention and possibly a few stitches. You would say, 'Alexa, ask Northwell what's the shortest wait time near 11021?' Alexa would then query the database of our emergency and urgent care wait times, which are refreshed every 15 minutes, and look at the locations nearest to the Zip code 11021."

Of those locations, it would then calculate the shortest wait time and report back the location's name, the wait time and the address. Say the result that comes back is for a Northwell GoHealth Urgent Care location, but the patient would prefer to go to the emergency department at their preferred hospital. In this case, the patient could for personal care attendant schedules. The technology can then check on schedules to see who is going to be on the next shift. They can set reminders for routine tasks such as taking medications, setting up doctor appointments and other medical events.

A Commonwealth Care Alliance member would commonly use voice assistant technology to set up a calendar or schedule for his or her personal care attendant. A typical exchange might take place as follows:

Patient: Alexa, create an event: PCA No. 1 at 9 a.m. to 12 p.m. on Monday.

Alexa: Does this repeat?

Patient: Yes.

At this time, the personal care attendant shift is entered into the Google calendar. Once the Google calendar is complete, it is shared with the personal care attendant so the staff member is aware of when the shifts are and who he or she may contact to arrange for coverage if shifts are missed.

The patient then can ask Alexa, "What are my next five events?" Alexa would inform the patient who is scheduled for his or her upcoming shifts. This is a powerful use of voice technology because the patients don't have to keep asking others to check the usually written schedules.

LIBERTANA: VOICE TECH HELPING INDEPENDENT LIVING

Libertana Home Health is using Alexa along with the Amazon Echo Dot. Its goal is to empower people to live independently.

Fig. News article about voice assistants

Methodology

In this project, we have used Google Cloud Speech to Text API to convert Medical Conversations to text and apply NLP techniques to generate Electronic Health Record. The dataset contains audio files from various sources containing medical conversations and dictations.

Cloud Speech-to-Text provides three methods for speech recognition:

1. Synchronous- Requests are limited to audio data of 1 minute or less in duration.
2. Asynchronous- Initiates a long running operation which offers requests for upto 480 minutes of audio length. We have used Asynchronous requests for our project.
3. Streaming- Captures live audio from a microphone.

Fig. Speech-to-Text

KDD Diagram:

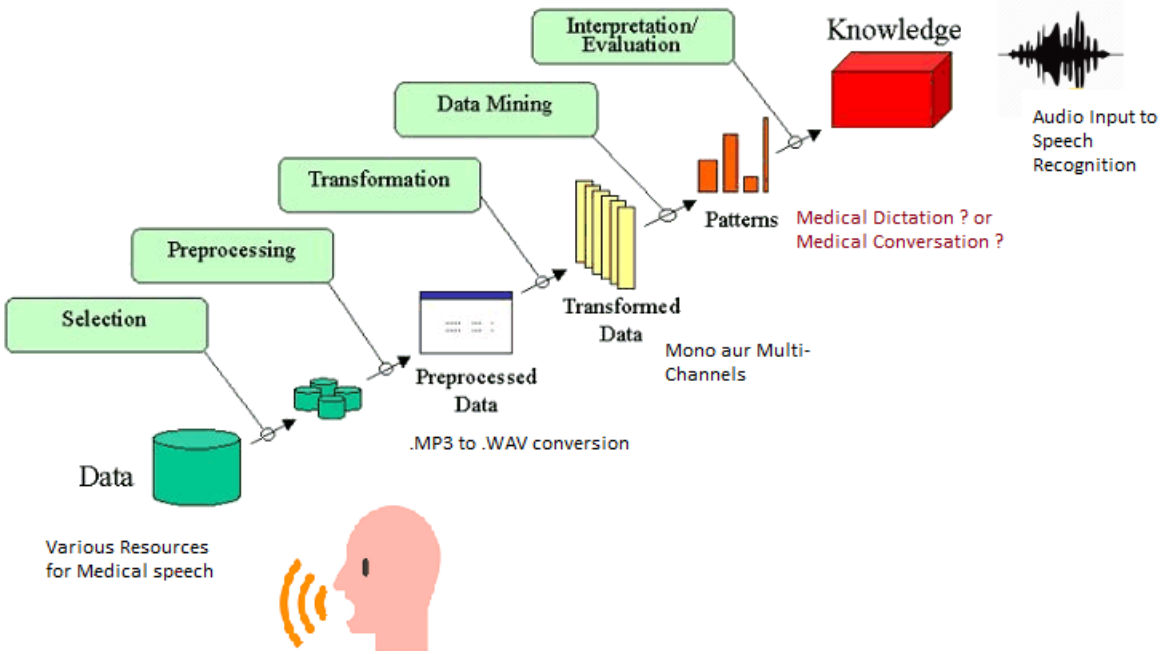

Fig. KDD diagram

Feature Engineering

- Out of the three main methods given by the speech recognition platform, we have decided to use the Asynchronous Recognition as the audio files required Long Running operation and were more than 1 minutes of duration.
- Audio file requirements needed to be converted from .MP3 to .WAV
- Model required us to specify sample rate of our audio in the sampleRateHertz, which only accepted rates between 8000Hz and 48000Hz.
- For audios with conversation, audio contained two channels, we needed to use multiple channels in such case and specify audio_channel_count.

The Data Bucket

The Data bucket contains all our audio files combined from various resources, which contains either Medical conversations or Medical Dictations.

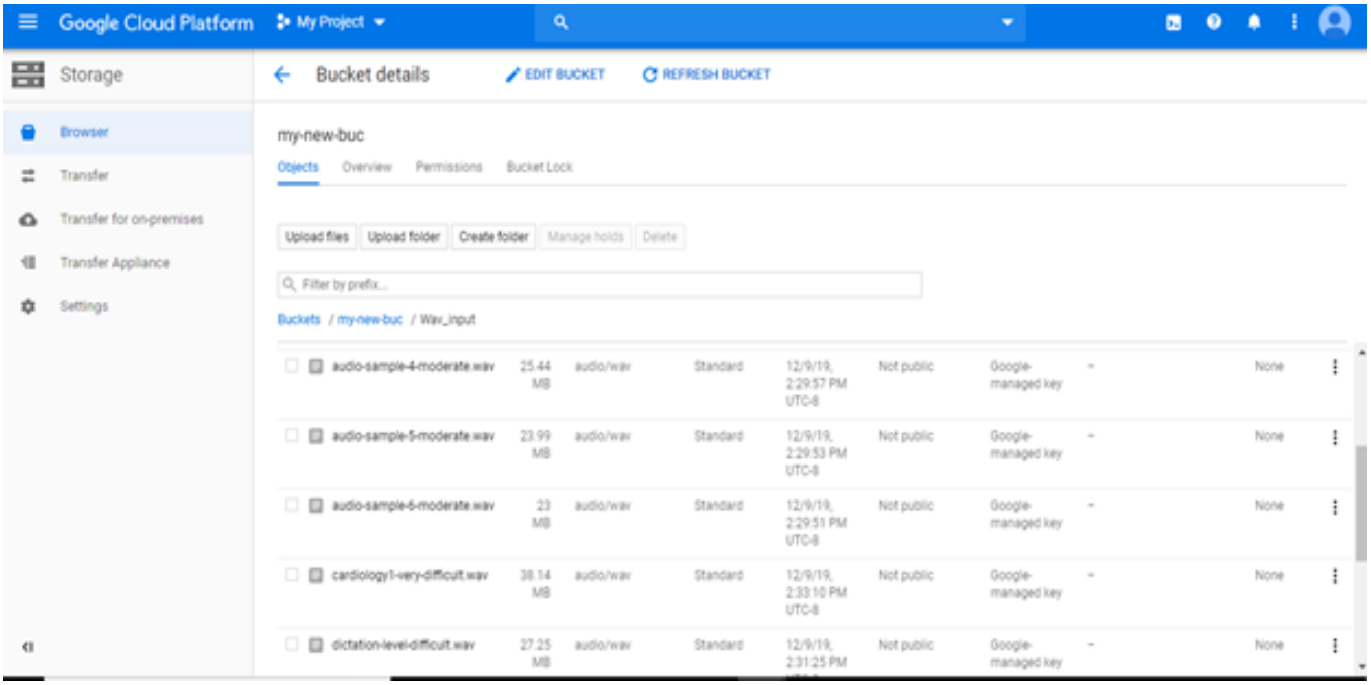

Fig. Google Data Storage Bucket

High Level Architecture Design

The Medical audios are selected and featured according to the requirements of the speech recognition model and are fed to the data bucket. The Google Cloud Platform program then fetches this dataset with the use of Natural Language Client Library. The transcribed text is evaluated based on the two evaluation methods and then pre-processed for the Electronic Healthcare Record (EHR).

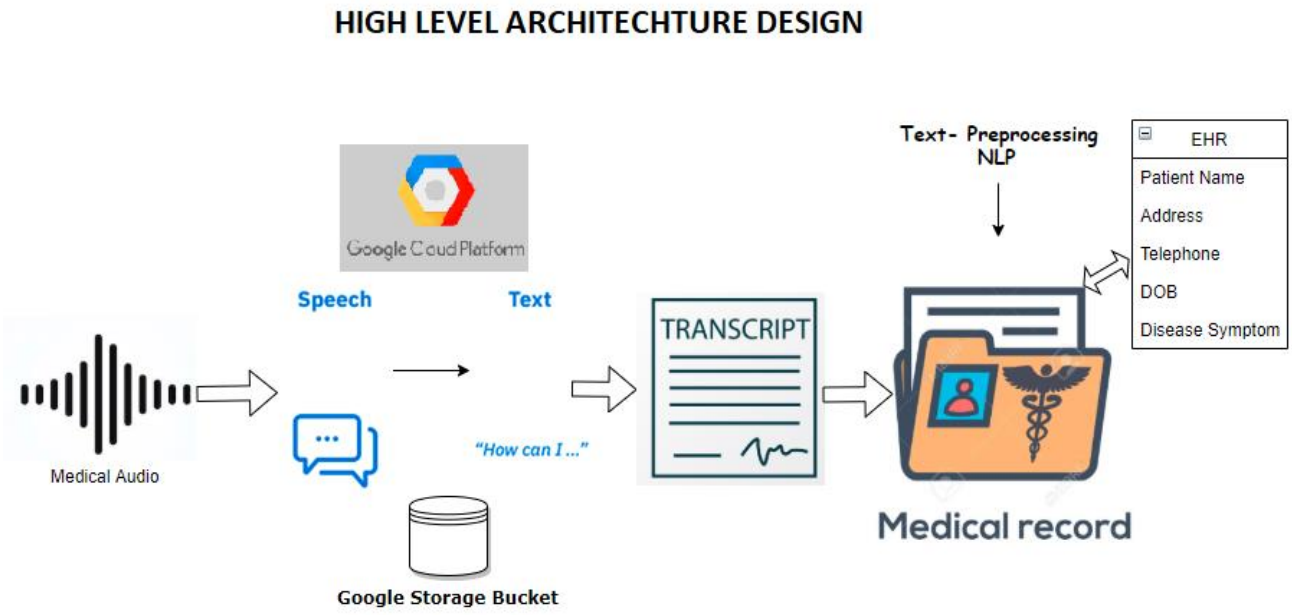

Fig. Our Architecture Design

Data Flow Diagram & Component Level Design

The figure below shows how the Asynchronous Speech Recognition works.

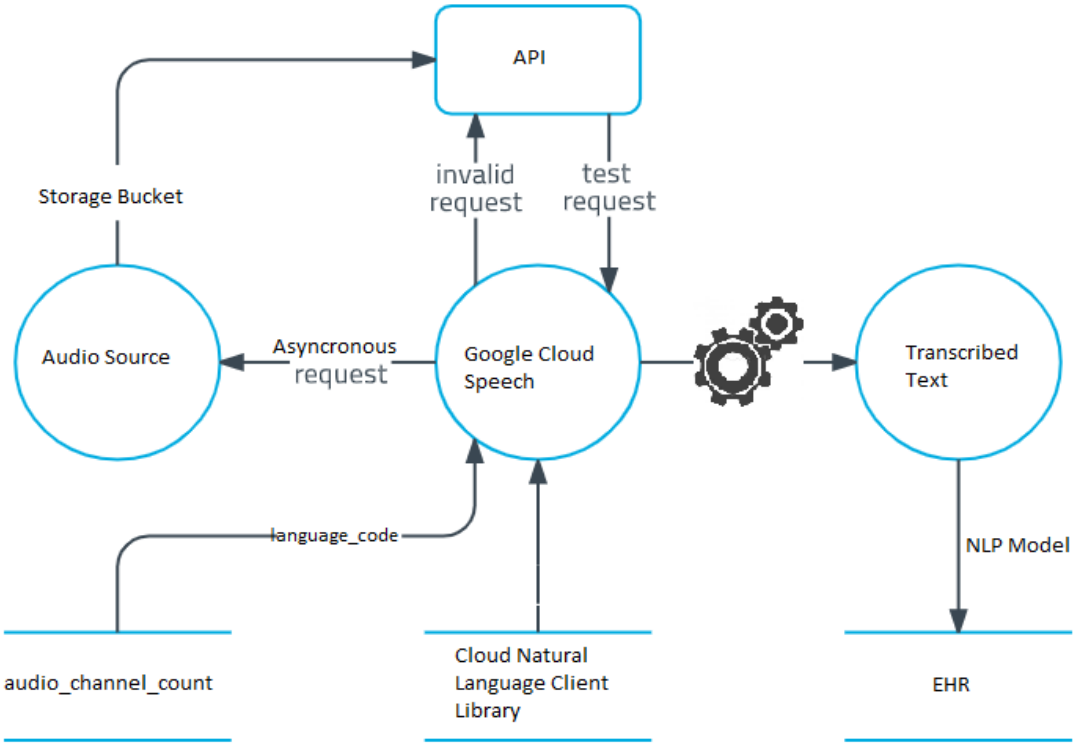

Fig. Dataflow Diagram.

Google Speech to Text API

Google speech to text is doing extremely easy to deploy, its free and just need to set up the environment and set the API key getting from google cloud and connected to the machine. Google speech to text transcription can convert both single channel and double channel audio file. The accuracy of it is around 95%.

Google Maps API

We are using Google Places API from Google Maps to detect the unstructured address from translation. All we need to do is set API key and enter project details and we're set to go. The main advantage of using this API is that we do not need to correct the structure manually. Moreover, it even corrects the misinterpreted text and returns a structured address in the form of a string to generate the pdf report.

Text Preprocessing and NLP

SpaCy, BIO Formatting:

In this article, authors are using the SpaCy library for extracting the entities. SpaCy is an open-source library for advanced Natural Language Processing in Python. It is designed specifically for production use and helps build applications that process and "understand" large volumes of text. Tokenization, Parts-of-Speech (PoS) Tagging, Text Classification and Named Entity Recognition are provided by SpaCy.

BIO stands for Beginning, Inside and Outside (of a text segment). In a system that recognizes entity boundaries only, only three labels are used: B, I and O while in a NERC system, entity classes are encoded into beginning and inside labels. Below is an example of BIO scheme with entity classes.

TAG	DESCRIPTION
B EGIN	The first token of a multi-token entity.
I N	An inner token of a multi-token entity.
L AST	The final token of a multi-token entity.
U NIT	A single-token entity.
O UT	A non-entity token.

Fig. BIO scheme

BIO tag structure:
[(TargetString_1, Tag, Label),
(TargetString_2, Tag, Label),
(TargetString_3, Tag, Label),...(TargetString_N, Tag, Label)].
the authors used library en_core_web_sm and the label of Date and Person to extract all the date and name information from the text file.

TF-IDF and Visualization

Wordcloud of the transcribed text is below, words like patient, pain, today, examination, opinion are the most frequently occurring terms.

Fig. Wordcloud

Results and Evaluation

Below is the Electronic Heath Record we have generated after performing all the NLP techniques to the transcribed text from the audios. We are able to achieve quite a good amount of accuracy. The health record contains patients name, address, telephone, date of birth and the symptoms of disease has been detected. We have then evaluated our speech recognition model using two methods, first is the word error rate (WER) and second is the Confidence.

Welcome to EHR!

Name: jason

Tel: 7146884899

Address:1 Washington Square, San Jose, CA 95112, USA

Date: last year 7146884899 date birth january 21st 1992 next monday wednesday week

DOB: january 21st 1992

Symptoms(Detected): face skin

Fig. Electronic Health Record from Medical Conversation

Word error rate

Word error rate is an industry standard to measure model accuracy. It counts the number of incorrect words identified during a speech recognition and then divides it by the total number of words in a reference transcript which is typically a human-led transcript. This number is then multiplied by 100%.

Word Error Rate can be calculated as follows-

$$WER = \frac{I + D + S}{N} * 100\%$$

Where,

I = incorrectly added words in transcript

D= undetected words in the transcript

S= words substituted between reference transcript and the speech recognition transcript.

Doctor's Assistant: Good morning, Doctor Franky's office. How may I help you?
Patient: Hello, I'd like to make an appointment to see Doctor Franky, please.

Doctor's Assistant: Have you been in to see Doctor Franky before?
Patient: Yes, I have. I had a physical last year.

Doctor's Assistant: Fine, what is your name?
Patient: Jason.

Doctor's Assistant: what's is your phone number?

Fig. Reference Transcript

```
Transcript: good morning talk to you Frankie's office how many I help you
hello I like to make appointment to see dr. Frankie face
Transcript: good morning talk to you Frankie's office how many I help you
hello I like to make appointment to see doctor Frankie please
Transcript: have you been to see doctor Frankie before yes I have I had
a physical last year by what's your name Jason what's your phone number 7
```

Fig. Speech-to-text output transcript

From the above output, we can observe that;

I = talk, to, you, by (4)

D = Doctor, please, an, in (4)

S = face, fine (2)

WER = 4+4+2 / 51 * 100 = **19.6%**

A WER of 5%-10% is considered to be good quality and is ready to use. A WER of upto 20% is still acceptable, however you may want to consider additional training. This WER can be considered for this transcript.

Confidence

Confidence score is an evaluation given by the Google Speech to text which is an estimated value between 0.0 and 1.0. It is a degree of accuracy for the entire transcription. Google cloud also provides confidence for individual word. This measure is calculated by aggregating the likelihood values assigned to each word in the given audio. Estimated greater likelihood that the individual words are recognized correctly is indicated by a higher number. It is noted that the first alternative is the most probable result.

Below are the Confidence scores for our medical audio-

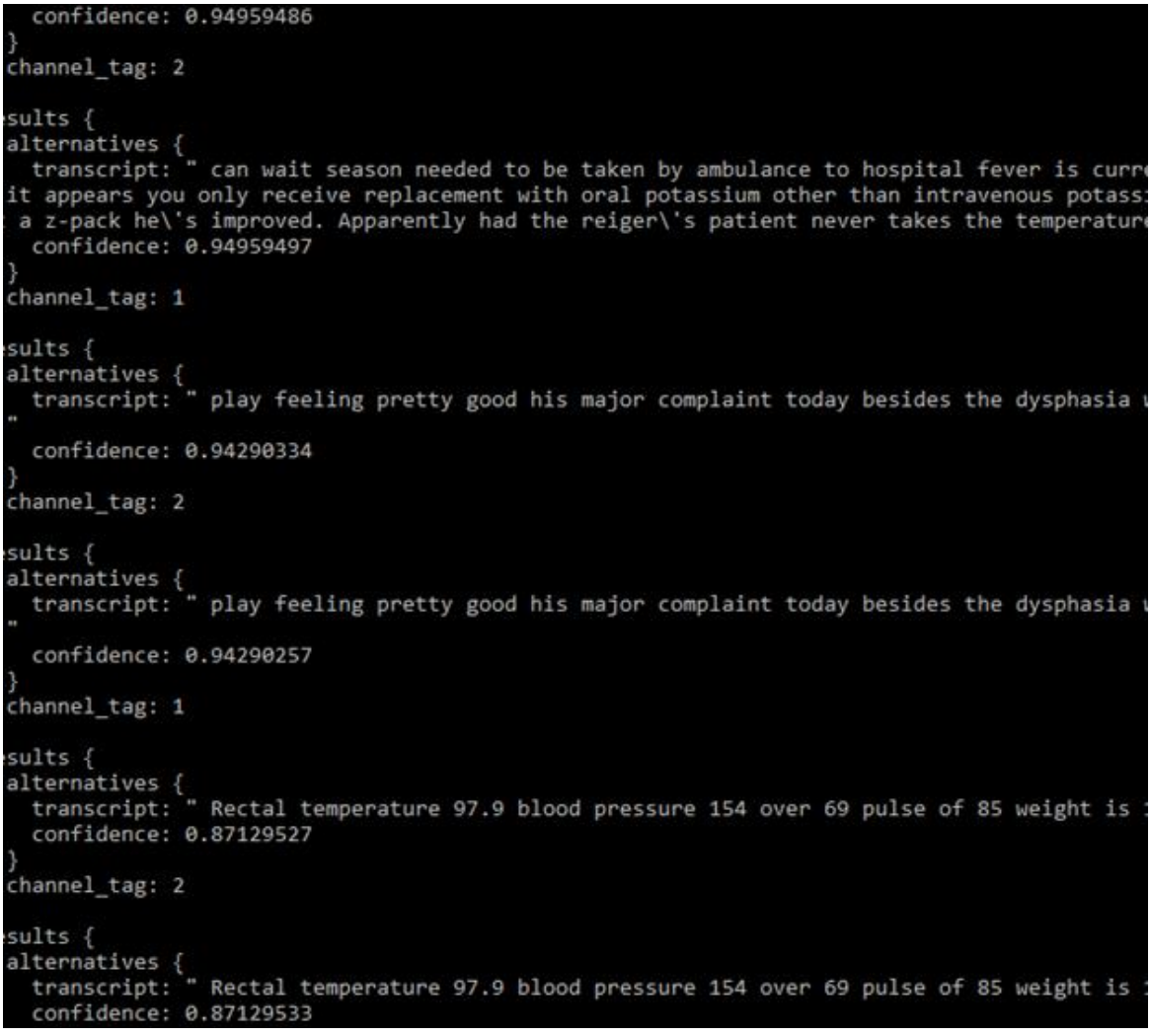

Fig. Transcribed text output with confidence

Conclusion and Future Work

We have successfully generated a health record based on the medical conversation between doctors and patients. This Electronic Health Record comes very handy and useful to the doctors. Medical Transcription is a big Industry, which is totally relying on the human-led transcriptions right now. Building and creating new and accurate models can help overcome the drawbacks prevailing with the existing methods. Our EHR model can be extended for detecting patient's diseases based on the conversation between doctor and patients with the use of existing API's and different Speech Recognition Models.

The Project Rubric

Requirements	• Setup a Google Cloud Console Service Account Project • Download Natural Language API • Set Credentials
Data Science Algorithms & Features Used	TFIDF, Text Preprocessing, NLTK, SpaCy, Speech Recognition
Interfaces –RESTFul& Server SideDesign	Google Cloud Console
Client SideDesign	Javascript for Google Maps
Testing (Data Validation / nFold)	as in feature engineering
Model Deployment	Python program using NLP API
HPC	We have used the Google Cloud Platform to console to integrate our data with the Google Speech to Text API.
Documentation	Project Report, PPT, Readme and code attached
Design Patterns Used	Synchronous, Asynchronous and Streaming Speech Recognition
AutoML or Serverless AI	Google Auto ML, Azure Cognitive Services
Data Engineering	Differentiating between Medical Dictation and Medical conversations
Active Learning or Feedback loop	EHR for detecting patient's information and diseases.
Interpretability of the Model	Created an EHR from the audio files.

References

[1] http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html

[2] http://transcription411.com/dictationsamples.htm

[3] https://corpora.uni-hamburg.de/hzsk/de/islandora/object/spoken-corpus:demo-1.0#corpus-content

[4] https://spacy.io

[5] https://s3.amazonaws.com/pageturnpro2.com/Publications/201803/3570/83350/PDF/1316441420385400 00_131638833965088000_131637804920963434_HITN_0318_1_40_optimized1.pdf

[6] https://developers.google.com/maps/documentation

[7] https://www.businessnewsdaily.com/8847-best-medical-transcription-services.html

[8] https://nnlm.gov/mcr/training/technology-program/electronic-health-records-ehrs

[9] https://www.barnesandnoble.com/w/the-big-book-of-nlp-expanded-shlomo-vaknin/1129948826