

A Helping Hand for The Visually Challenged, Mute and Deaf people (Team – 6)

Assignment - 2

Submitted by

RAMKRITHIK S (18BCE0528)

S A HARIPRASAD (18BCE0686)

N KISHORE KUMAR (18BCE2106)

KUNAL AGARWAL (18BCE2108)

A ANIRUDDHA (18BCI0029)

PARDHASARADHI TIRUMALASETTI (18BCE2068)

Prepared For

TECHNICAL ANSWERS FOR REAL WORLD PROBLEMS (CSE3999)

TC1 Slot

Submitted To

Dr. ARCHANA T

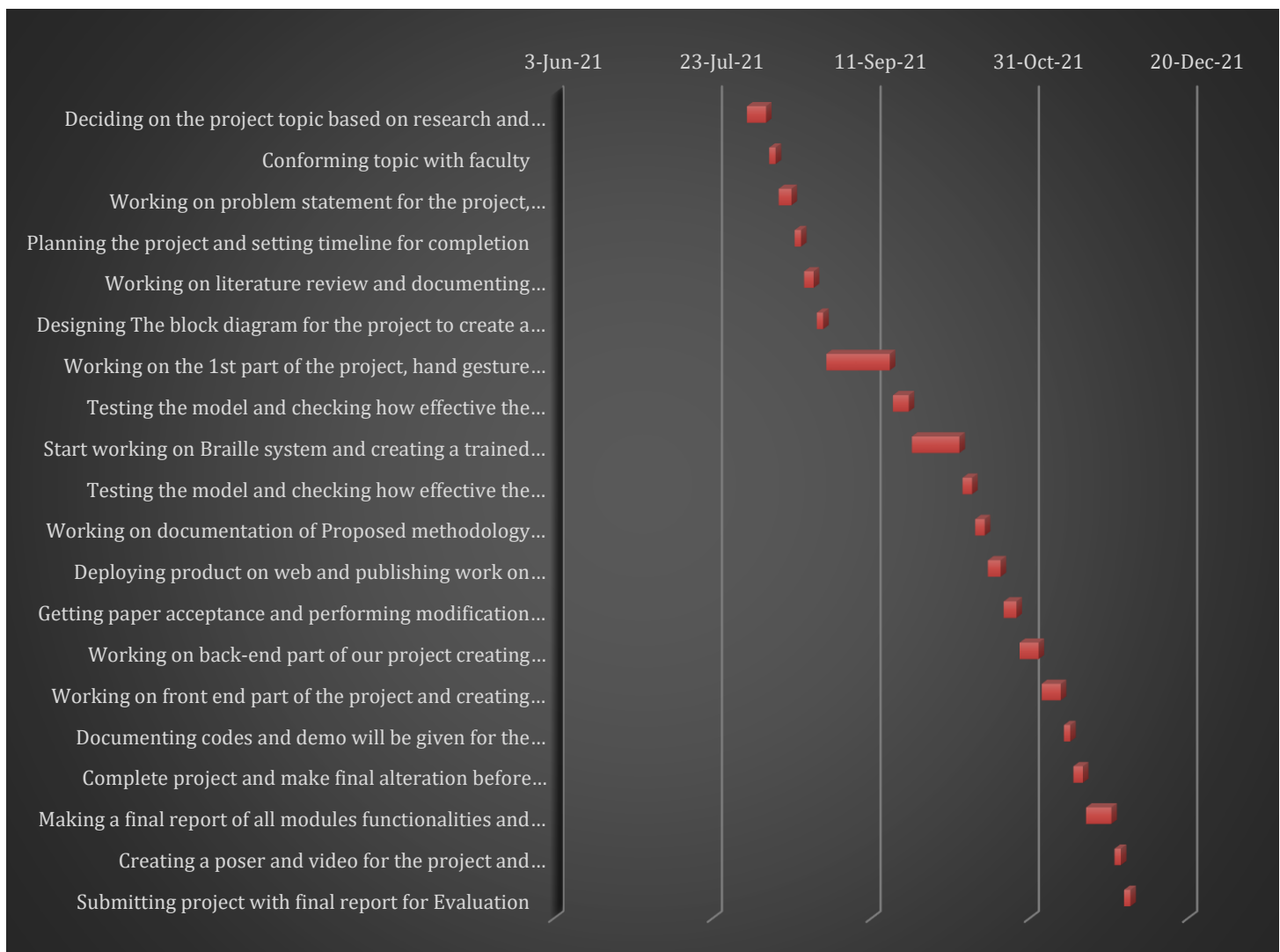
School of Computer Science and Engineering (SCOPE)



1. PROJECT PLAN:

TASK ID	TASK DESCRIPTION	TASK DURATION	START DATE	END DATE
1	Deciding on the project topic based on research and study of recent trends	6	3-Aug-21	9-Aug-21
2	Conforming topic with faculty	2	10-Aug-21	12-Aug-21
3	Working on problem statement for the project, deciding on project outcome and giving a detail explanation about the project	4	13-Aug-21	17-Aug-21
4	Planning the project and setting timeline for completion	2	18-Aug-21	20-Aug-21
5	Working on literature review and documenting existing survey	3	21-Aug-21	24-Aug-21
6	Designing The block diagram for the project to create a basic understanding of the project	2	25-Aug-21	27-Aug-21
7	Working on the 1st part of the project, hand gesture recognition developing a trained model	20	28-Aug-21	17-Sep-21
8	Testing the model and checking how effective the model works on command line	5	18-Sep-21	23-Sep-21
9	Start working on Braille system and creating a trained machine learning model	15	24-Sep-21	9-Oct-21
10	Testing the model and checking how effective the model works on command line	3	10-Oct-21	13-Oct-21
11	Working on documentation of Proposed methodology and producing detailed description of work carried out	3	14-Oct-21	17-Oct-21
12	Deploying product on web and publishing work on reputed journal.	4	18-Oct-21	22-Oct-21
13	Getting paper acceptance and performing modification based on requirements.	4	23-Oct-21	27-Oct-21
14	Working on back-end part of our project creating routes and functionalities	6	28-Oct-21	03-Nov-21
15	Working on front end part of the project and creating UI for the application	6	04-Nov-21	10-Oct-21
16	Documenting codes and demo will be given for the product	2	11-Nov-21	13-Nov-21

17	Complete project and make final alteration before deployment	3	14-Nov-21	17-Nov-21
18	Making a final report of all modules functionalities and documenting all works	8	18-Nov-21	26-Nov-21
19	Creating a poser and video for the project and submitting all documents	2	27-Nov-21	29-Nov-21
20	Submitting project with final report for Evaluation	2	30-Nov-21	2-Dec-21
	Total	102		



[Click here to view Full image](#)

2. INTRODUCTION:

With progresses in new technologies, cell phones have filled in notoriety to become perhaps the most well-known customer device. Phones are vital piece of current life. Large numbers of us need to settle on a decision or communicate something specific at whenever from anyplace. For visually impaired and motion-impaired individuals this issue is more self-evident, however others likewise regularly deal with this issue, e.g., when driving or utilizing a PDA under splendid daylight. Located clients frequently discover them unavoidably put under circumstances where non-visual interaction is required.

Individuals have been confronting a ton of trouble in regards to different issue, particularly incapacitated individuals in which Blindness is one of the most, if not the most, misconceived sort of impairment. The general masses have their own pre-envisioned ideas about the visually impaired individuals that they determinedly acknowledge to be substantial without contacting a visually impaired person. We will make a talk bot for mute and deaf individuals, for mute individuals we will make a sign language converter, which will change sound over to sign language and show it. For the visually impaired individuals we will do a Braille converter which will get a text or a picture from the client and will change that over to Braille code which will be useful for blind individuals. We'll trust that our website will actually want to help individuals experiencing the scourges of this world so they will not surrender and continue to hustle. We will utilize React for the front end.

The visually impaired individuals face difficulties day by day in speaking with their general surroundings. They need to rely upon their located associates for settling on a telephone decision and getting to other portable functionalities. This framework is a voice perceiving application for cell phones that permit admittance to the vast majority of the functionalities of the telephone and will make it feasible for visually hindered individuals to associate with the general public. The located client's kin with restricted perusing capacity can likewise utilize our website/application who are knowledgeable with the Braille language.

Additionally, our point is to work on the correspondence with individuals who has hearing hardships and utilizing any sign language to articulate their thoughts. At the principal sight, as a

thought, how troublesome could make a sign dialects converter. After nitty gritty examination about sign language linguistics, it is sorted out around 240 sign dialects have exist for communicated in dialects on the planet.

We'll catch capacities and specialized provisions to the motion catch of sign to voice Change. The phonology and Morphology can be clarified as the phonology alludes to the investigation of actual sounds present in human discourse. The phonology of sign language can be characterized. Rather than sounds, the phonemes are considered as the unique signs present in succession of hand signs.

They are considering the accompanying boundaries:

1. Configuration: Hand shape while doing the sign.
2. Orientation of the hand: Where the palm is highlighting.
3. Position: Where the sign is finished.
4. Motion: Movement of the hand while doing the sign (straight, influencing, circularly) [2].
5. Contact point: Which portion of the hand contact the body.
6. Plane: The sign is relying upon the distance to the body.
7. Non-manual parts: Information given by the body. For instance, when the body inclines front, it communicates future tense.

Morphology:

Communicated in dialects have inflectional morphology and furthermore derivational morphology. The inflectional morphology alludes to the change of words. The derivational morphology is the cycle of shaping another word based on a current word. Sign dialects have as it were derivational morphology in light of the fact that there are no infusions for tense, number or individual. The most significant boundaries viewing morphology are addressed as:

1. Degree: Mouthing.
2. Reduplication: Repeating a similar sign a few times.
3. Compounds: Fusion of two distinct words.
4. Verbal Aspect: Expressing action words in an unexpected way. A few of these include reduplication.
5. Verbal number: To communicate plural or solitary action words. Reduplication is additionally used to communicate it.

3. LITERATURE REVIEW:

1. Research on the Hand Gesture Recognition Based on Deep Learning

Authors: Jing-Hao Sun, Ting-Ting Ji, Shu-Bin Zhang, Jia-Kui Yang, Guang-Rong Ji

Published Date: 6 Dec. 2018

Publisher: IEEE

With the quick improvement of PC vision, the interest for communication among human and machine is turning out to be increasingly broad. Since hand signals can communicate advanced data, the hand motion acknowledgment is generally utilized in robot control, shrewd furnishings and different viewpoints. The paper understands the division of hand signals by building up the skin shading model and AdaBoost classifier dependent on haar as indicated by the distinction of skin tone for hand motions, just as the denaturation of hand motions with one edge of video being cut for investigation. In such manner, the human hand is divided from the convoluted foundation, the continuous hand motion following is additionally acknowledged by Cam Shift calculation. Then, at that point, the space of hand signals which has been identified continuously is perceived by convolutional neural organization to understand the acknowledgment of 10 normal digits. there are numerous approaches to acknowledge hand motion division. In light of the division of skin shading model, the skin shading model is set up to understand the hand motion division as per the contrast between skin shade of hand signals and outer climate and the model isn't influenced by the hand stances, however it can't prohibit the articles which are like the skin tone, like human face, etc; the hand motion division dependent nervous discovery can section the hand motions as per the brokenness of Gray worth in the edge space of picture locale, yet it is not difficult to be hindered by the commotion and it has severe prerequisites for the foundation; the hand motion division dependent on development data, including outline contrast technique and foundation distinction strategy, etc. takes on the data of development of hand signals to section hand motions on the reason of static of foundation. The impact is acceptable in static climate while not all around acted in dynamic foundation; the division strategy for hand signal dependent on measurable format coordinating can quickly distinguish the hand region and non-hand region by utilizing preparing classifier of motion layout include, yet it can just remember at least one hand motions, it can't fulfill our requests. The hand motion division in the paper pre-measures the pictures and sets up

Gaussian combination model as per the skin tones, additionally, it likewise portions hand signals by joining with AdaBoost classifier dependent on Haar highlights.

Pros and Cons: Hand motion division by utilizing AdaBoost classifier dependent on Haar highlight understands the procurement of hand signal region in muddled climate. Utilizing CamShift calculation for hand motion following as per the development of hand motions and components of twisting guarantees to secure the hand signal region continuously, at long last, the hand motion region is ordered by convolution neural organization.

2. End-to-End Attention-based Large Vocabulary Speech Recognition

Authors: Dzmitry Bahdanau, Jan Chorowski, Dmitriy Serdyuk, Phil'emon Brakel, Yoshua Bengio

Published Date: 19 May 2016

Publisher: IEEE

The framework proposes here is a neural organization that can plan groupings of discourse edges to arrangements of characters. While the entire framework is differentiable and can be prepared straightforwardly to play out the main job, it can in any case be separated into various useful parts that cooperate to figure out how to encode the discourse signal into an appropriate component portrayal and to unravel this portrayal into a succession of characters.

Pros and Cons: The subsequent methodology is altogether less difficult than the overwhelming HMM-DNN one, with less preparing stages, many less helper information and less area ability included. Joined with a trigram language model our framework shows respectable, albeit not yet best in class execution.

3. SQuAD: 100,000+ Questions for Machine Comprehension of Text

Authors: Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang

Published Date: 16 Jun 2016

Publisher: IEEE

System: Candidate answer age. For each of the four strategies, as opposed to thinking about all $O(L^2)$ ranges as competitor replies, where L is the quantity of words in the sentence, we just use traverses which are constituents in the voting demographic parse created by Stanford CoreNLP. Disregarding accentuation and articles, we track down that 77.3% of the right replies in the improvement set are constituents. This places a successful roof on the exactness of our strategies. During preparing, when the right reply of a model is certainly not a constituent, we utilize the briefest constituent containing the right reply as the objective. Sliding Window Baseline For every up-and-comer reply, we figure the unigram bigram cross-over between the sentence containing it (barring the actual applicant) and the inquiry. We keep every one of the competitors that have the maximal cross-over. Among these, we select the best one utilizing the sliding-window approach. Calculated Regression In our strategic relapse model, we extricate a few sorts of components for every up-and-comer reply. We discretize each consistent element into 10 similarly estimated cans, fabricating a sum of 180 million provisions, the majority of which are lexicalized components or reliance tree way includes.

Pros and Cons: Goal of regular language understanding, we present the Stanford Question Answering Dataset, an enormous perusing appreciation dataset on Wikipedia articles with publicly supported question-answer sets. Execution separated by syntactic difference. As talked about in Section 4, one more testing part of the dataset is the syntactic difference between the inquiry and answer sentence. shows that the more difference there is, the lower the exhibition of the calculated relapse model. Strangely, people don't appear to be delicate to syntactic dissimilarity, proposing that profound agreement isn't diverted by shallow contrasts.

4. Attention Is All You Need

Authors: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin.

Published Date: 12 Jun 2017

Publisher: IEEE

Most cutthroat neural grouping transduction models have an encoder-decoder structure. Here, the encoder maps an info grouping of image portrayals to an arrangement of nonstop portrayals $z =$

($z_1; \dots; z_n$). Given z , the decoder then, at that point produces a yield succession ($y_1; \dots; y_m$) of images each component in turn. At each progression the model is auto-backward, devouring the recently created images as extra information while producing the following. The Transformer follows this general engineering utilizing stacked self-consideration and point-wise, completely associated layers for both the encoder and decoder.

Pros and Cons: To assess if the Transformer can sum up to different assignments, we performed investigates English supporters parsing. This errand presents explicit difficulties: the yield is dependent upon solid underlying requirements and is altogether more than the information. Moreover, RNN succession to-grouping models have not had the option to achieve cutting edge brings about little information systems. the main succession transduction model dependent on consideration, supplanting the intermittent layers most generally utilized in encoder-decoder structures with multi-headed self-consideration.

5. Adaptive Document Retrieval for Deep Question Answering

Authors: Bernhard Kratzwald, Stefan Feuerriegel

Published Date: 20 Aug 2018

Publisher: IEEE

System: Threshold-Based Retrieval As a credulous pattern, we propose a basic edge-based heuristic. That is, not set in stone with the end goal that the total certainty score arrives at a decent edge. Ordinal Regression-further carry out a teachable classifier as an ordinal edge relapse which is custom fitted to positioning assignments. It is additionally expected that total certainty liable to be direct. The classifier then, at that point approximates n_i with an expectation y_i that signifies the situation of the primary important report containing the ideal reply.

Pros and Cons: Set up that profound inquiry noting is dependent upon a commotion data compromise. As an outcome, the quantity of chose records in profound QA ought not be treated as fixed, rather it should be painstakingly custom fitted to the QA task. Second, we propose versatile plans that decide the ideal report count. This can extensively support the exhibition of profound QA frameworks across different benchmarks. Third, we further exhibit how significant

a versatile archive recovery is with regards to various corpus sizes. Here our versatile technique presents an adaptable procedure that can effectively adjust to it and, contrasted with a proper report count.

6. Ranking Paragraphs for Improving Answer Recall in Open-Domain Question Answering

Authors: Jinhyuk Lee, Seongjun Yun, Hyunjae Kim, Miyoung Ko, Jaewoo Kang

Published Date: 1 Oct 2018

Publisher: IEEE

Open-space QA frameworks are built as pipelines that incorporate a recovery framework and a peruser model. We also constructed Paragraph Ranker that helps our QA pipeline for a superior section determination. For the recovery framework and the peruser model, we utilized Document Retriever also, Document Reader. Section Ranker-Answer Aggregation.

Pros and Cons: By utilizing Paragraph Ranker, the QA pipeline benefits from expanded answer review from sections to peruse, and channels insignificant reports or passages. With our straightforward Paragraph Ranker, we accomplish cutting edge exhibitions on the four open-area QA datasets with enormous edges. As future works, we intend to additionally further develop Paragraph Ranker dependent on the explores on figuring out how to rank.

7. BERT: Pre-preparing of Deep Bidirectional Transformers for Language Understanding

Authors: Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova

Published Date: 11 Oct 2018

Publisher: IEEE

There are two stages in this structure: pre-preparing and calibrating. During pre-preparing, the model is prepared on unlabeled information over various pre-preparing errands. For finetuning, the BERT model is first introduced with the pre-prepared boundaries, and the entirety of the

boundaries are tweaked utilizing marked information from the downstream undertakings. Each downstream errand has separate adjusted models, despite the fact that they are introduced with similar pre-prepared boundaries. Pre-preparing BERT-Fine-tuning BERT.

Pros and Cons: Recent exact enhancements because of move learning with language models have shown that rich, solo pre-preparing is a basic piece of numerous language getting frameworks. Specifically, these outcomes empower even low-asset undertakings to profit from profound unidirectional structures.

8. DocChat: An Information Retrieval Approach for Chatbot Engines Utilizing Unstructured Documents

Authors: Zhao Yan, Nan Duan, Junwei Bao, Peng Chen, Ming Zhou, Zhoujun Li, Jianshe Zhou

Published Date: 2016

Publisher: IEEE

Considering a much improved on task, short text discussion (STC) in which the reaction R is a short text and just relies upon the last client expression Q . Past strategies for the STC task generally depend on Q - R matches and fall into two classifications: Retrieval-based techniques (e.g., Ji et al., 2014). This sort of techniques initially recovers the most conceivable hQ^{\wedge}, Ri^{\wedge} pair from a bunch of existing Q - R sets, which best matches current expression Q dependent on semantic coordinating with models, then, at that point accept R^{\wedge} as the reaction R . One hindrance of such a strategy is that, for some, particular areas, gathering such QR sets is unmanageable. Age based strategies. This sort of techniques normally utilizes an encoder-decoder system which initially encode Q as a vector portrayal, then, at that point feed this portrayal to decoder to create reaction R . Like recovery-based strategies, such methodologies likewise rely upon existing Q - R sets as preparing information. Like other language age errands, like machine interpretation and rewording, the familiarity and naturality of machine produced text is another disadvantage. To defeat the issues referenced above, we present a clever reaction recovery approach, DocChat, to discover reactions dependent on unstructured archives. For every client expression, rather than searching for the best Q - R pair or creating a word arrangement dependent on language age strategies, our strategy chooses a sentence from given archives straightforwardly, by positioning all potential sentences

dependent on highlights planned at various degrees of granularity. On one hand, utilizing reports as opposed to Q-R matches enormously work on the adjust.

Pros and Cons: One inadequacy of the unaided subject model is that, the point size is pre-characterized, which probably won't mirror reality on a particular corpus. Reaction recovery can't generally assurance to return an up-and-comer set that contains somewhere around one appropriate reaction, yet reaction positioning will yield the most ideal competitor constantly. Thus, we need to choose which reactions are sufficiently sure to be yield, and which are not.

9. Optical Recognition of Braille Writing Using Standard Equipment

Authors: Jan Mennens, Luc van Tichelen, Guido Francois, and Jan J. Engelen

Published Date: Dec 1994

Publisher: IEEE

Using the property that Braille characters are constantly situated on a decent network, it first attempts to assemble a framework comprising of flat and vertical lines that go through every one of the spots, and afterward it checks in case there is a speck present on every one of the convergence focuses. The lattice development should be adaptable in light of the fact that there are situations where it very well may be disfigured or unpredictable. This relies incredibly upon the nature of the first.

Pros and Cons: Although Braille dabs are set on a decent framework, certain Braille creation procedures cause this network to be unpredictable. - Deformations of the Braille Cell-Deformation of the Grid Where Braille Characters are Positioned. The primary justification behind fostering a framework that can peruse Braille is to safeguard and increase enormous volumes of physically created books. Many books on math or music are undeniably challenging, in any event, for a talented copyist, to retype because of the uncommon principles that apply in Braille.

10. Teaching Machines to Read and Comprehend

Authors: Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, Lei Yu, and Phil Blunsom

Published Date: 10 Jun 2015

Publisher: IEEE

Distributed/neural models permits us to learn shallow components for our classifiers, catching basic connections between bits of feedbacks. Intermittent neural organizations give an exceptionally down to earth apparatus to sentence implanting. Review interpretation as encoding and unraveling sentences.

Pros and Cons: Twenty years prior log-direct models permitted more noteworthy opportunity to demonstrate relationships than basic multinomial parametrizations, yet forced the requirement for highlight designing. This venture kills the issue.

11. DRISHTI—A Gesture Controlled Text to Braille Converter

Authors: Vineeth Kartha, Dheeraj S. Nair, Sreekant S., Pranoy P. and Dr. P. Jayaprakash

Published Date: 28 January 2013

Publisher: IEEE

The device consists of a microcontroller which is the heart of the device. The conversion of characters take place in this device, at present we have started the work with Atmega 8 chip as it was available to us but the chip had to be changed to Atmega 328 due to memory constraints of Atmega 8. The code of Atmega 8 is compatible with Atmega 328, so switching the chip will not hinder the progress made so far. The tactile display is made of six solenoids that represent the Braille characters, the device will be having only a single Braille cell. Atmega328- Touchpad The microcontroller code using the Arduino platform. Software on the computer using processing platform, poppler utils and unoconv

Pros and Cons: Suggestion made by a person who tested this device was that by including upto 5 braille cells the speed problem can be solved. Also, since the device we made was a prototype it had imperfections in the leveling of the braille cell. use of piezoelectric Braille cell will further improve speed of the device and also reduce the power consumption. The present device is noisier due to the solenoids using piezo-electric braille cells will reduce the noise too. Another development possible is to find suitable batteries that can be used to power up the device so that it can be used even without AC mains. Integrating USB drives in place of Micro SD cards is also a possible expansion on the hardware side. On the software side using optical character recognizers opens a new arena for the visually challenged to learn and read. The Braille displays are superior to text to speech synthesizers as it promotes Braille literacy.

12. Development of a Text to Braille Interpreter for Printed Documents through Optical Image Processing

Authors: Joshua L. Dela Cruz, Jonaida Angela D. Ebreo, Reniel Allan John P. Inovejas, Angelica Romaine C. Medrano, and Argel A. Bandala

Published Date: 3 Dec. 2017

Publisher: IEEE

Haptic System, OCR, Image Processing, Text Conversion. - the camera will capture an image, in tiff format, which will be analyzed through an optical image processing integrated in a GUI. After analysis, a universal asynchronous receiver/transmitter (UART) device will interface the communication between the computer and the microcontroller. The microcontroller then sends the data to the refreshable braille cells. Camera- Optical Image Processing- USB to UART Device- Microcontroller- Braille Cells.

Pros and Cons: The solution presented by this study is better than the similar and existing studies because the development of this system is definitely more modernized or smart compared to the existing ones. This system will give the visually-impaired people the chance to read any printed reading paraphernalia. The system is able to achieve an overall system speed of 1 word in 2 seconds, and an overall system reliability of 95.68%. It is easy to use but is not stand-alone. The

system is efficient both in character recognition and in actuation of braille cells, making it a good alternative to braille printed books.

13. Albert: A Lite Bert for Self-Supervised learning of Language Representation.

Authors: Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, Radu Soricut

Published Date: 9 Feb. 2017

Publisher: ICLR

Expanding model size while pertaining natural language portrayals frequently brings about further developed execution on downstream performance. Nonetheless, eventually further model increments become more harder because of GPU memory constraints and longer training times. To resolve these issues, this paper presents two parameter reduction methods to bring down memory utilization and increment the training speed of BERT. They additionally utilized a self-supervised loss that spotlights on modeling between sentence coherence. There are some other solutions to solve to aforementioned problems which are model parallelization and clever memory management, these solutions address the memory limitation issue but not the communication overhead. This Albert incorporates two parameter reduction method that lift the communication overhead.

Pros and Cons: This paper has overcome the issue with huge models by using two parameter reductions with cross layer parameter sharing which has shown promising results. But the data used for pertaining is the same size as the original BERT and the batch size is also similar. The drawback of using the BERT model is that they are huge and significantly limit their adoption.

14. RoBERTa: A Robustly Optimized BERT Pretraining Approach

Authors: Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov

Published Date: 26 July. 2019

Publisher: ICLR

Training is computationally costly, regularly done on private datasets of various sizes, and, as we will show, hyperparameter decisions fundamentally affect the final outcomes. This paper presents a replication investigation of BERT pertaining that carefully measures the effect of many key hyperparameters and preparing information size. We find that BERT was fundamentally undertrained, and can coordinate or then again surpass the presentation of each model published after it. These outcomes feature the significance of recently neglected plan decisions, furthermore, bring up issues about the wellspring of as of late detailed upgrades. The results in this paper have shown that dynamic masking is slightly better than static masking which was then used for the remainder of the experiment in the paper.

Pros and cons: The data used for pretraining is nearly 10 times more than the data the original BERT uses which produces accurate results. It is also trained with a batch size eight times larger for half as many optimization steps. The drawback of using Bert model is that they cannot run on normal machines and require massive resources to even get them running properly

15. Improving Language Understanding by Generative Pre-Training

Authors: Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever

Published Date: 2017

Publisher: IEEE

Although huge unlabeled text corpora are plentiful, labeled information for learning these particular tasks is scarce, making it hard for discriminatively prepared models to perform satisfactorily. We exhibit that enormous gains on these tasks can be acknowledged by generative pre-training of a language model on a different corpus of unlabeled text, followed by discriminative tweaking on each explicit task. Rather than past approaches, we utilize task-aware input changes during fine-tuning to accomplish successful exchange while requiring minimal changes to the model design. The paper shows the effectiveness of the methodology on a wide scope of benchmarks for regular language understanding. The overall undertaking skeptic model

beats discriminatively prepared models that use models specifically created for each task, essentially refining the best in class in 9 out of the 12 assignments studied.

Pros and cons: In this paper the main transfer was to capture higher level semantics while previous were word-level information. In recent work, the method has been used to help train deep neural networks on various tasks.

16. MURAX: A Robust Linguistic Approach For Question Answering Using An On Line Encyclopedia

Authors: Julian Kupiec

Published Date: 01 July 1993

Publisher: IEEE

Robust semantic techniques are applied to the assignment of addressing shut class questions utilizing a corpus of normal language. The techniques are shown in a wide space: addressing general-information questions utilizing an on-line reference book. A shut class question is an inquiry expressed in regular language, which expects some unmistakable answer embodied by a thing expression instead of a procedural reply. The techniques theorize thing phrases that are probably going to be the appropriate response, and present the client with pertinent message where they are checked, focusing the client's consideration suitably. Moreover, the sentences of coordinating with message that are displayed to the client are chosen to affirm express relations suggested by the inquiry, instead of being chosen exclusively based on word recurrence. The corpus is gotten to through a data recovery (IR) framework that upholds Boolean hunt with closeness requirements. Inquiries are consequently built from the phrasal substance of the inquiry, and passed to the IR framework to discover applicable text. Then, at that point the applicable text is itself broke down; thing phrase speculations are extricated and new questions are freely made to affirm state relations for the different theories.

Pros and Cons: Task assessment demonstrates where further exertion may be generally useful and furthermore shows how new parts add to by and large execution. The semantic investigation depends on a hidden normal language formalism, both in the HMM tagger and the expression

perceives. There might be benefits from the utilization of stochastic setting free language structures, which can likewise be prepared from unlabeled text and empower uncertainty to be evaluated in probabilistic terms.

17. Part-of-Speech Tagger for Biomedical Domain Using Deep Neural Network Architecture

Authors: A. Gopalakrishnan, K. P. Soman and B. Premjith

Published Date: 2019

Publisher: ICCCNT

In the given paper the author has developed a pos tagger which helps in the process of classifying words into part of speech like nouns, verbs, prepositions, etc. The pos tagger helps in speech recognition, sentiment analysis, etc. Here the developer has developed deep learning models like RNN, LSTM, and GRU and check which will give better performance. The dataset used is taken in GENIA. The main reason for developing this neural network model is because the previous model the hidden Markov model(hmm) which is a statistical model which has an emission matrix and a transition matrix produces an accuracy between 70-80% only and even while trying with a hybrid model the accuracy was only increased by 2-3%. The proposed model the developer develops the model using RNN, LSTM, bi-directional RNN, bi-directional LSTM, and also gated recurrent unit. After the training of the model, the bi-directional LSTM produced the highest accuracy of 94.8% which is better than the hidden Markov model.

Pros and Cons: This accuracy is very good in the biomedical field because in that field the number of sequences is very high and developing models using this accuracy provides better results. Hence the use of neural networks has enhanced the accuracy of the POS tagger. The only future enhancement that can be achieved is using a better GPU to reduce the time taken to develop the model.

18. Polish Language Modelling Based on Deep Learning Methods and Techniques

Authors: Piotr Kłosowski

Published Date: 18 Sept, 2019

Publisher: IEEE

This paper presents an example of deep learning methods application for language modelling in polish. Developing a language model helps to predict a sequence of recognized words or characters that can be helped and useful to make speech recognition. This is done using a neural network model as it provides better accuracy than a statistical model. The deep learning model helps in achieving a sophisticated word-based and character-based model. It is a hybrid model containing a recurrent neural network and deep machine learning technique. This hybrid model helps to characterized by giving better properties and also improves speech recognition. The paper provides results for the polish language but by changing the dataset similar results can be achieved for other languages also.

Pros and Cons: Through the given hybrid model the developer was able to get an accuracy of 93.6% for a word-based type and an accuracy of the character-based model for a character-based model which is 10-15% better than a machine learning model. The biggest problem is as the training set was very big the time taken to train this model came more than 30-40 hours. In the Future scope, the developers can try to reduce the time taken by better parallelization and improving GPU performance of the system in which the model is trained. Hence the developer was able to make a very good hybrid model that has helped to improve the speech recognition of the polish language

19. Electronic Medical Data Analysis Based on Word Vector and Deep Learning Model

Authors: L. Li, C. Niu, D. Pu and X. Jin

Published Date: 2018

Publisher: ITME

This paper presents an electronic medical data analysis based on word vector and deep learning. This has been developed because of the continuous promotion of medical data processing and the

electronic data field is rapidly developing. This model was mainly developed to minimize the problem caused by a large amount of redundancy, missing semantic, and ambiguity in the text parts of electronic medical data. The vectorized data is combined with the LSTM model and Yoon model. The above model is combined with comprehensive machine learning models and has shown significant results in the text classification of electronic medical data. After combining the deep learning model with machine learning models like SVM(support vector mechanism), random forest, naïve Bayes, decision Tree, and KNN(k mean clustering) the SVM has shown the highest accuracy of 83.1% and KNN showed the lowest accuracy of 69.2%.

Pros and Cons: The test results show that using hybrid machine learning and deep learning model has not only improved the accuracy but also has improved the precision and recall of the model. The problem faced by previous statistical and pure machine learning model has been overcome. In the future, the developers can try to improve the performance of the model by adding more layers for the deep learning model and also better pre-processing of the data.

20. Deep Learning Techniques for Part of Speech Tagging by Natural Language Processing

Authors: Rushali Deshmukh, Arvind Kiwelekar

Published Date: March, 2020

Publisher: ICIMIA

The author and her team have developed a part of speech tagger using deep learning techniques for a widely spoken popular language Marathi. As Part of speech tagging is considered very important for the development of natural language processing applications and also the previous machine learning models were not showing great results a deep learning model was developed to make the model. In this paper, the developers have made 3 fold deep learning model and a bi-directional long short term memory to perform pos tagging for Marathi text. For the given model the dataset used is a corpus containing 1500 sentences from Marathi e-newspaper using 32 tags. In the early research works performed using machine learning less than 32 tags were considered. The developer has assumed each sentence is assumed to be of length 100 and a size 100 vector is

used to represent every word. The usual 20% of data is used for testing and 80% for training. After developing both the models the 3-fold machine learning model showed an accuracy of close to 80% and the bi-LSTM model showed an accuracy of close to 97%.

Pros and Cons: The deep learning model accuracy was able to out beat the machine learning model by more than 15-20%. The model has also given scope for other Indian languages to be developed using this POS tagger model. In future works, the developers can increase the size of the corpus to develop a stronger and more precise model.

21. Some Experiences on Applying Deep Learning to Speech Signal and Natural Language Processing

Authors: Y. Liao and Y. Wang

Published Date: 2018

Publisher: DISA

In the given paper the author yuan-fu Liao has developed different deep learning techniques to improve speech signals in natural language processing. In this paper, deep learning techniques are applied to sentiment analysis, microphone array speech, speech recognition, language recognition, and text-to-speech concepts of natural language processing. The different types of deep learning techniques used are convolution neural networks, recurrent neural networks, and long short-term memory. To develop these models the datasets used are taken from the university sources. In the speech enhancement, an RNN model is developed where 4 hidden layers are used. After successfully developing the model, it has been observed that the accuracy is showing close to 85% which is better than the machine learning model. Then the sentiment analysis was developed using a bi-LSTM model using a 2-skip gram. It was observed that the accuracy for this sentiment analysis was close to 72% which is much better than the machine learning model which showed an accuracy of 48%. Then an audio event detection was developed using a convolution neural network. This has had a very high accuracy of 98% which is 12% more than a machine learning model. Then a speech recognition system was developed using a multi-lingual deep learning model. The accuracy of the model was very good compared to the machine learning model which showed an accuracy of close to 62%. The speech recognition and speech synthesis was also developed using a DNN

based text-to-speech system which showed an accuracy of 72% which showed an accuracy of 67% which is better and more promising compared to the machine learning models.

Pros and Cons: This paper has clearly shown the significance of the use of deep learning techniques for the development of speech signals compared to normal machine learning techniques. In future enhancement, the developers can try to automate this process using semi-supervised and supervised training algorithms.

22. Dzongkha Word Segmentation using Deep Learning

Authors: Y. Jamtsho and P. Muneesawang

Published Date: 2020

Publisher: KST

In the given paper the author Yeshe jamtsho has developed a word segmentation model for word segmentation of Dzongkha (Bhutan language) using deep learning. This language was not considered because the language is very complex and also the sentence is written as a string of syllables without proper word boundaries. The main novelty in this work is that it has used deep learning techniques to achieve this avoiding manual feature engineering. The windows approach was also incorporated where the syllable tag depends on the surrounding. The dataset was a private Bhutan dataset containing 9203 sentences where each word ends with begging. The deep learning model which has been developed has 3 layers like an embedded layer, a flattened layer, and a linear layer, and then the results are normalized and sent through an activation function which finally returns the tag for word segmentation. The experiment was performed with a pre-trained and non-pretrained dataset with a hyperparameter ranging from 0-4.

Pros and Cons: The accuracy was high with hyperparameter 3 and a pre-trained dataset giving an accuracy close to 95% as there are no previous research works performed in this field this is a very good accuracy for a system which has been developed for the first time. In future works, the developers can try to develop the model with a much bigger corpus and also improve new tags for single-syllable words. The butane language is filled with morphological and abbreviated words

which can be considered in future researches. Hence a new beginning has opened for more Bhutan based natural language processing models

23. Understanding Legal Documents: Classification of Rhetorical Role of Sentences Using Deep Learning and Natural Language Processing

Authors: S. R. Ahmad, D. Harris and I. Sahibzada

Published Date: 2020

Publisher: IEEE

In the given paper the author Rameel Ahmad has created a legal document understander by using the classification of the rhetorical role of sentences using deep learning and natural language processing in the year 2020. The main aim of this paper is to automatically detect patterns in legal documents and increasing on-time case processing. In this paper, a deep learning model is created to break down these legal documents and classify all the rhetorical sentences. This model will eliminate the problems faced through current exists through various legal systems. The configuration used in training the deep learning model is a bidirectional long short-term memory with 300-dimensional Glove embedding's which was also used previously for sentiment understanding of law documents. The architecture developed for this system contains an embedding layer, a bi-directional LSTM a fully connected layer a re-lu layer then a 0.5 dropout is done and a fully connected layer is fixed to a SoftMax. The model used a dataset from an open-source repository titled veteran claims.

Pros and Cons: After training and testing the data it was observed that the model was able to provide an accuracy of 76% and a precision of 74% which is much better than the machine learning model which only gave an accuracy of 52%. In future enhancements, the model can use a better pre-processing phase to improve the accuracy and precision of the data. This model will provide a great foundation for processing in many documentation fields and reducing the time taken for checking the legal documents for banks and many other sectors.

24. An Effective Natural Language Understanding Model using Deep Learning and PyDial Toolkit

Authors: K. Ganesan and A. P. Patil

Published Date: 2017

Publisher: IEEE

In the year 2017, the author Karthik Ganesan has developed a natural language understanding model using deep learning and PyDial Toolkit. In a world where everyone wants to understand any type of spoken system so to achieve that an effective natural language understanding model is required. For developing an efficient natural language understanding model, the developer has used a hybrid bi-directional LSTM and a recurrent neural network for identification. The dataset used is an AITS airline dataset that describes the slot under different categories.

Pros and Cons: The biggest advantage of developing this model is it will be able to understand any kind of language ranging from small context and also understanding multi-word types. This gives a scope for also building a spoken dialogue system in any domain. The main advantage of this system is very easy use and can be helpful in building basic voice-based search agents for any database or corpora. After training the data with the model and testing the data it has been observed that the accuracy of the model came close to 86% which is a lot better compared to the machine learning model which only gave an accuracy of 60%. It was also observed that the use of the neural network model with the pydial tool kit has enhanced the integration of multiple natural language processing forming an AI model that will help the users to conveniently use this natural language understander. In future enhancement, a new python module like pydial can be used, and also GPU of the system can be improved to process this big dataset.

25. Algorithm selection for classification problems

Authors: N. Pise and P. Kulkarni

Published Date: 1 September 2016

Publisher: IEEE

In this paper, algorithm selection is proposed for classification problems in data mining. The characteristics of datasets and the performance of classification algorithms are found out. Then based on the problem of classification, or the new problem at hand, mapping is done between the datasets and the benchmark performance of different classifiers. K-similar datasets are returned. Then ranking of classification algorithms is performed and based on the highest rank, the classification algorithm is recommended for the problem at hand. Hence the user doesn't need to waste time for working on different data mining algorithms, fine tuning the parameters for different algorithms. The algorithm is directly recommended for his problem.

Pros and Cons: The learning parameters used for training classifiers affects the performance of the classifiers. So, the further work is required to discover the suitable values for a given dataset. Grid search and evolutionary algorithms can be used for this task. This will be helpful in model selection which includes not only algorithm selection but also parameter optimization.

26. Hierarchical Anomaly Detection Model for In-Vehicle Networks Using Machine Learning Algorithms

Authors: Seunghyun Park and Jin-Young Choi

Published Date: July 2020

Publisher: MDPI

Methodology: Proposed a multi-labelled hierarchical classification (MLHC) intrusion detection model that analyzes and detects external attacks caused by message injection. This model quickly determines the occurrence of attacks and classifies the attack using only existing classified attack data. The proposed model can classify both the type and existence or absence of attacks with high accuracy and can be used in interior communication environments of high-speed vehicles with a high throughput. The analysis model is trained by injecting training data, and the performance of the trained model is evaluated using test data. The intrusion detection module, including the trained model in an actual application environment, is used to detect follow-up information, such as attack or benign, vehicle type and attack type, after receiving the CAN (Controller Area Network) message frame as input.

Pros and Cons: Thus, we conclude that the DT and RF algorithms are applicable to high-speed internal communication environments, as well as in CAN for analysing 43 million and 46 million CAN message frames per second, respectively. The disadvantage with this model is that, there can be loss of information and can lead to lots of classes with small number of data.

27. Efficient Distributed Preprocessing Model for Machine Learning-Based Anomaly Detection over Large-Scale Cybersecurity Datasets

Authors: Xavier Larriva-Novo ,Mario Vega-Barbas, Víctor A. Villagrà, Diego Rivera, Manuel Álvarez-Campana and Julio Berrocal

Published Date: May 2020

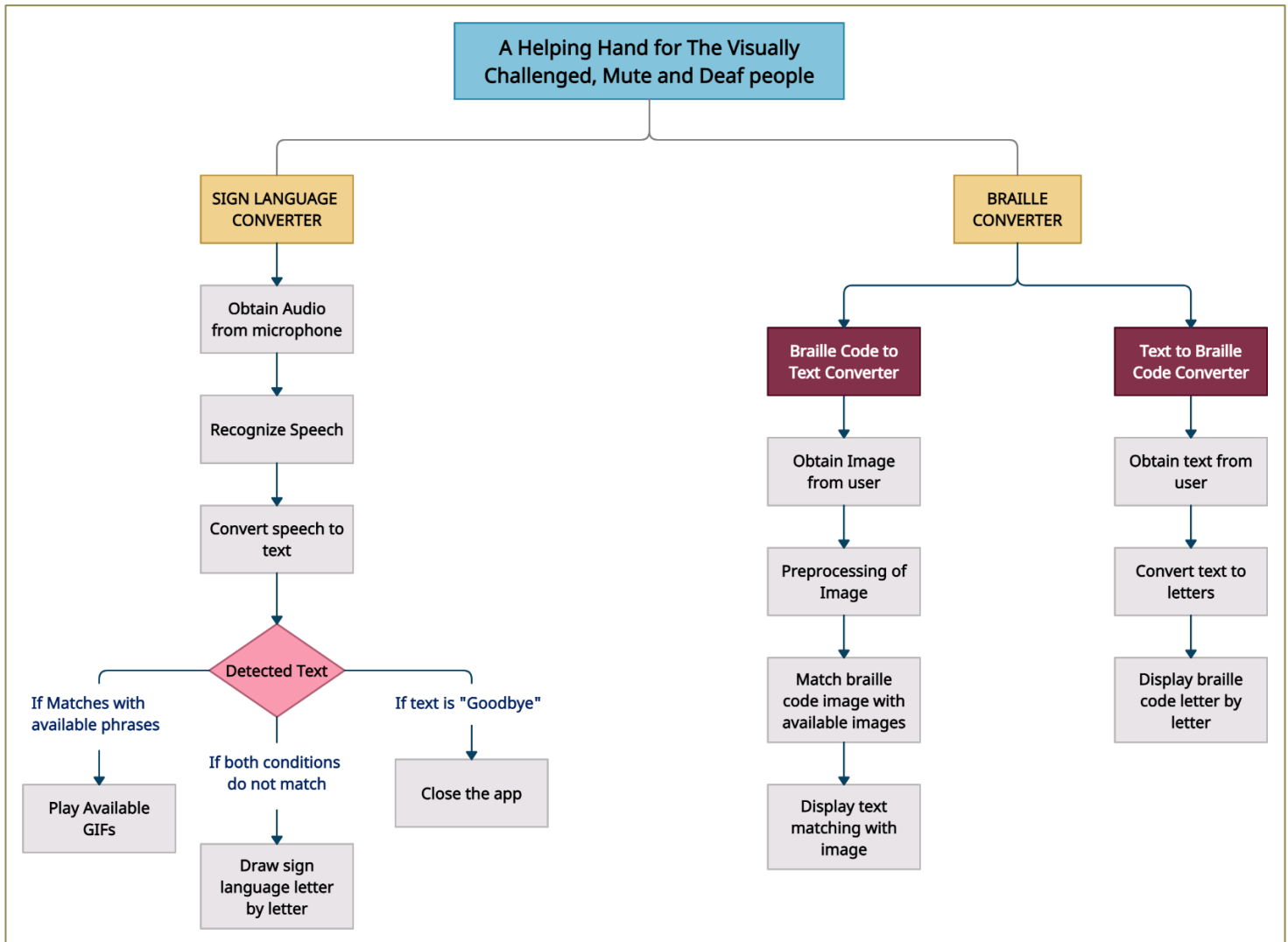
Publisher: MDPI

Methodology: This paper presents a new model of data pre-processing based on a novel distributed computing architecture focused on large-scale datasets. In addition to that, the paper analyses the use of machine learning techniques in order to improve the response and efficiency of the proposed pre-processing model. Methodology used here is, firstly select the most outstanding features of the dataset to process them correctly by the ML algorithm. Once the pre-processing is finished, the ML model training process should be performed. The pre-processing operation requires a large number of resources to be carried out correctly, which is something to consider carefully when designing it. The main objective of this research is presenting a new model to pre-process data in an optimized way, performing computer execution in time for large-scale datasets.

Pros and Cons: Using this model required a high amount of memory in order to train the pre-processed data. That memory can be up to five times the size of the dataset. The CPU can process this huge amount of data despite the high time needed to process it. Instead, by using GPUs, it is possible to obtain better results in terms of time due to the use of several GPUs at the same time. However, to perform training on several GPUs at once, it is necessary to add an overhead of data to the batches that, in each epoch, are sent to each GPU to relocate the results to the complete original model. This slows down the process due to the large overhead added.

4. OVERVIEW OF THE PROPOSED WORK:

4.1. Block Diagram:

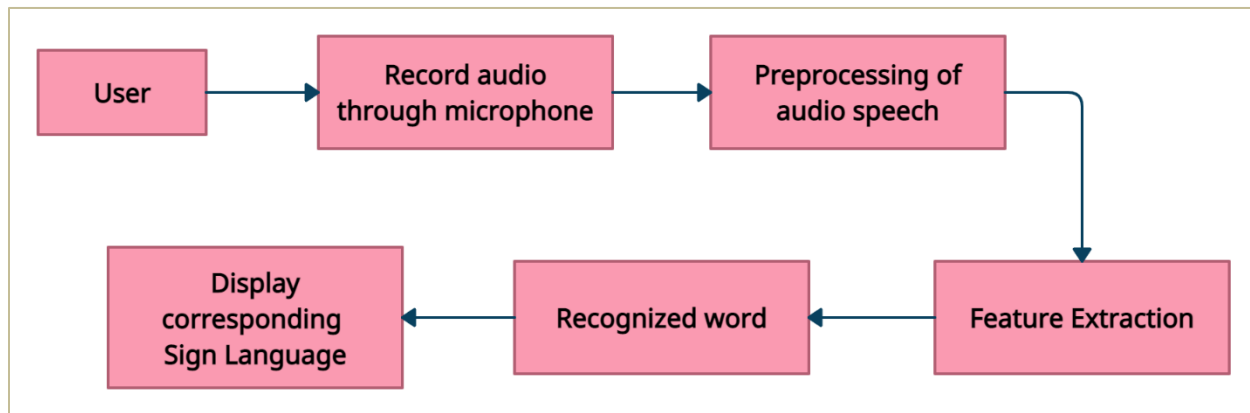


[Click here to view Full image](#)

We are creating sign language converter and braille converter system for our project. In sign language converter, the user will record the speech using microphone. With the help of speech recognition system, we will convert the audio to text form. If the detected text matches the, available phrases in database it will show the corresponding GIF, else if the text is “goodbye”, it will close the app, else if the text does not match with the available phrases, the system will display

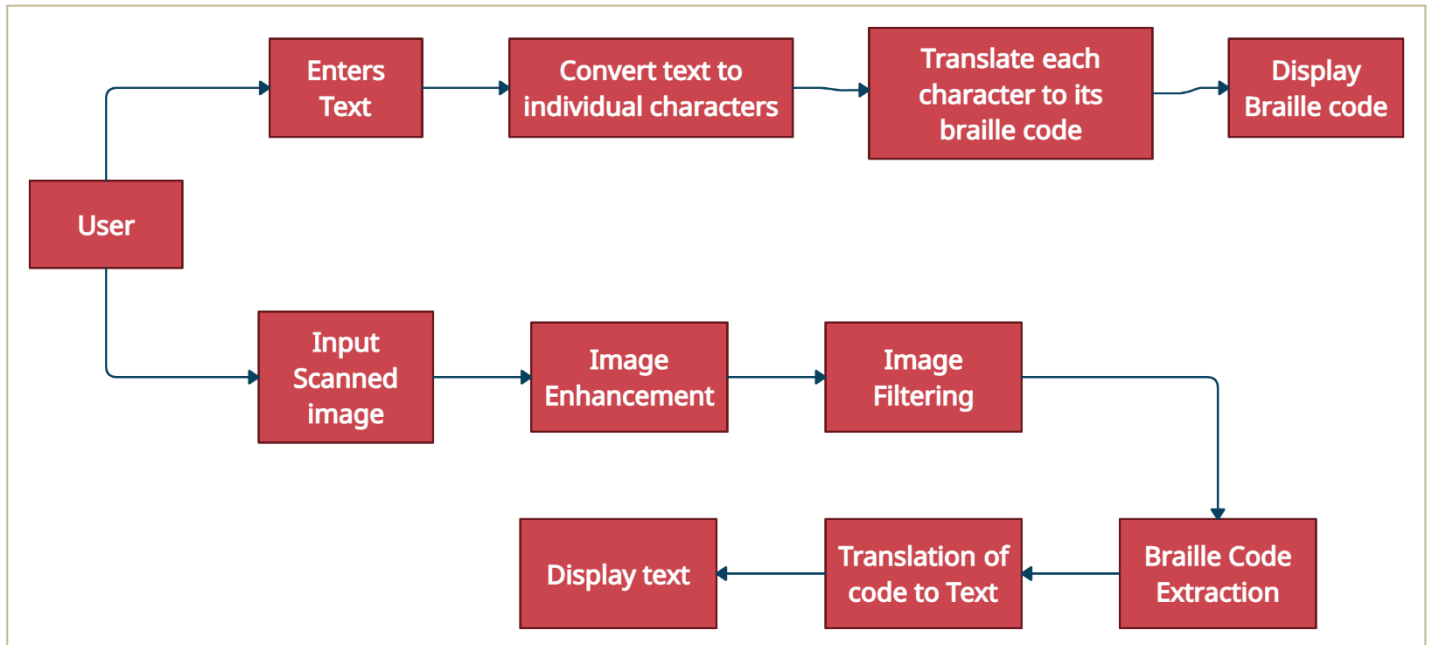
the sign language letter by letter. In braille system, there are two subparts. First, converting text to braille code, in which the user will enter input, the system will separate each letter and it will display the corresponding braille code. Second, converting braille code image to text, in which the user will input braille code image which will be preprocessed and matched with the available images in the database and print the corresponding text.

4.2. Data Flow Diagram – Sign language conversion:



In sign language converter, the user will record the speech using microphone. With the help of speech recognition system, we will convert the audio to text form. The speech recognition system will first pre-process the audio to enhance its quality. After pre-processing, the features are extracted from the audio using which the audio speech is classified into words which results in text. If the detected text matches the, available phrases in database it will show the corresponding GIF, else if the text is “goodbye”, it will close the app, else if the text does not match with the available phrases, the system will display the sign language letter by letter.

4.3. Data Flow Diagram – Braille System:



The user can either convert braille code to text, in which the user should input a braille code. The system will perform image enhancement and image filtering techniques to separate the braille codes into various segments. Then each segment of image is extracted and compared with the available images and its corresponding text is displayed to the user. User can also convert text to braille code in which the user should enter a text. The system will separate the text into letters, and each letter is translated into its corresponding braille code and displayed to the user.