VIRTUAL ASSISTANT WITH HAND GESTURE INPUT

Karthik Krishnan O(20BDS0104) , NithiyaSri M(20BCI0230), Karanam Sree HarshaVardhan(20BDS0243), Kumar Satyam(20BCE2383)

Vellore Institute of Technology, Vellore

Abstract

Virtual Assistants are very common in today's world. It is basically an application program that can replace a human assistant in performing basic tasks.These assistants help the visually impaired a lot but there are rarely any of these for deaf people. The project focuses on providing a platform for the deaf people to communicate with the computer using sign languages and get their work done. This can also help ordinary people to interact with these disabled people better than before. The project is an implementation of a voice+text+hand gesture based assistant which can play songs, movies, search the web , gives quick responses to the user's questions and a lot more like other voice assistants.

Keywords: `Virtual Assistant, Chat-bot, Hand gesture recognition, Voice Assistant, Music Player , Computer Vision, Speech Recognition, Text to Speech, Wake word detection`

1. INTRODUCTION

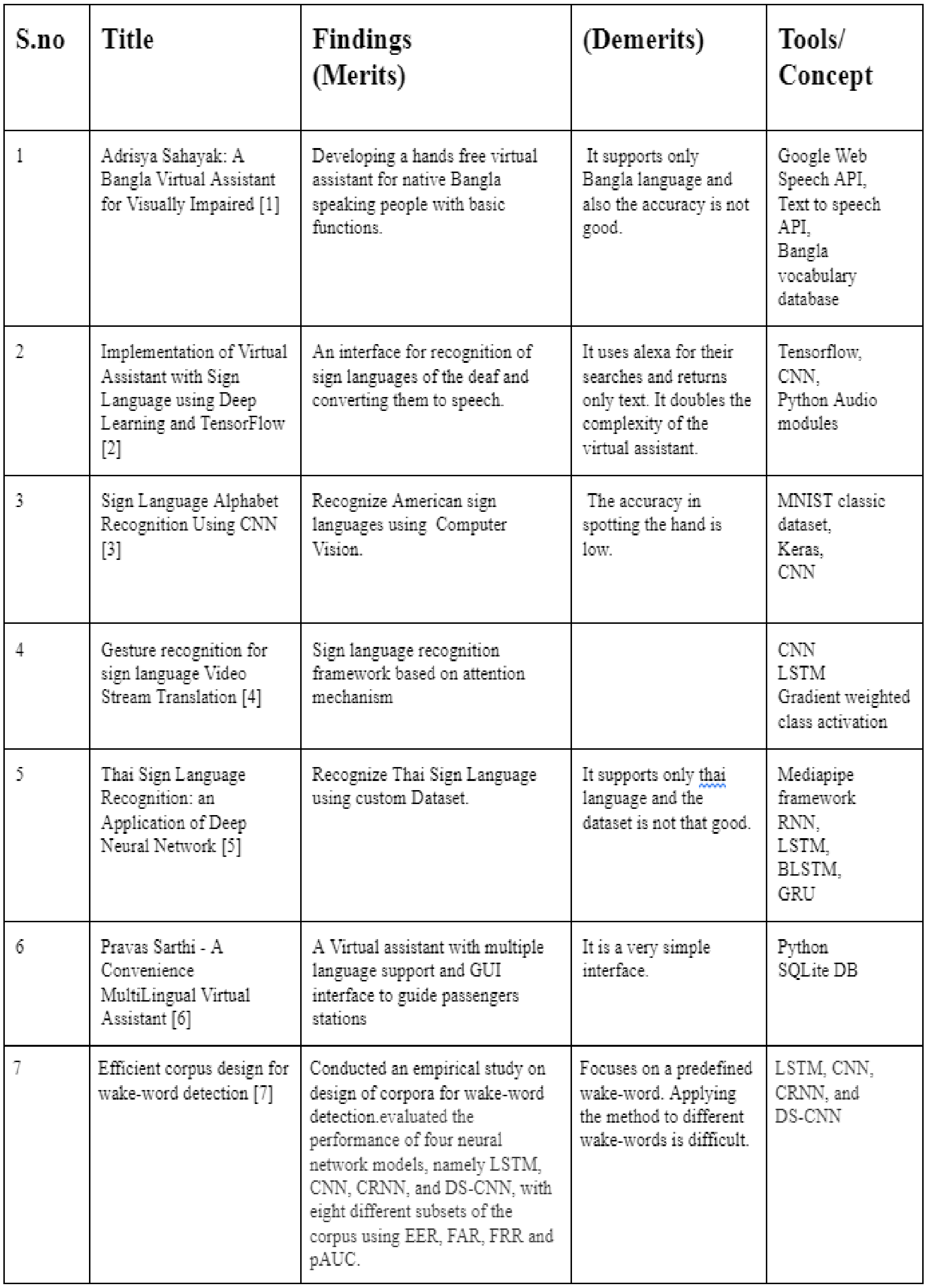
Virtual assistants are a member of every house and office. The tasks which are historically done by personal assistants are now replaced by virtual assistants. The modern technology allows even middle class or lower class people to have a virtual assistant in their pocket. The virtual assistant systems like the google assistant or Apple's Siri, Amazon's Alexa, Microsoft Cortana are now available in one's fingertips. They are available 24/7 to listen to the users commands and perform the tasks for them.

Most of the assistants available are voice based. This turns out to be a very good aid for the visually impaired people who finds it difficult to use digital keyboards. Till date there aren't any good application software for the deaf people. They can see but cannot hear or talk. They communicate with each other using sign languages which uses their hand most of the time.

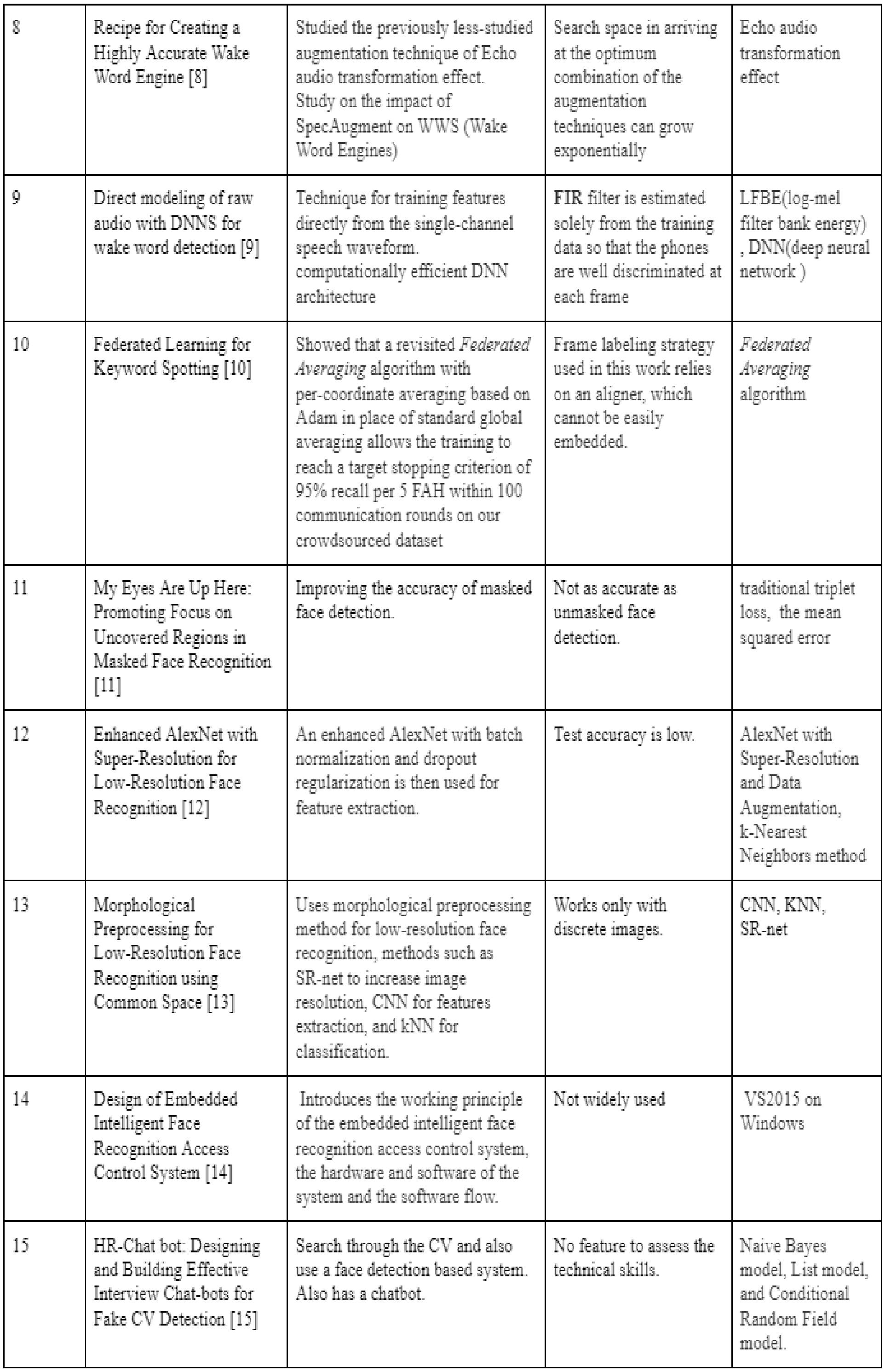
A personal assistant that can take the user's real time video as input, recognise the sign language and acts according to the user's need will be a very novel idea and also it will be a great achievement to bring up the deaf people in the society. Along with the hand gesture input, audio and text inputs are also available for the user's convenience.

Similar to other famous virtual assistants, this will also have a wake word or wake sign to run the program. The assistant will be on sleep mode for rest of the time which assures the privacy of the user.

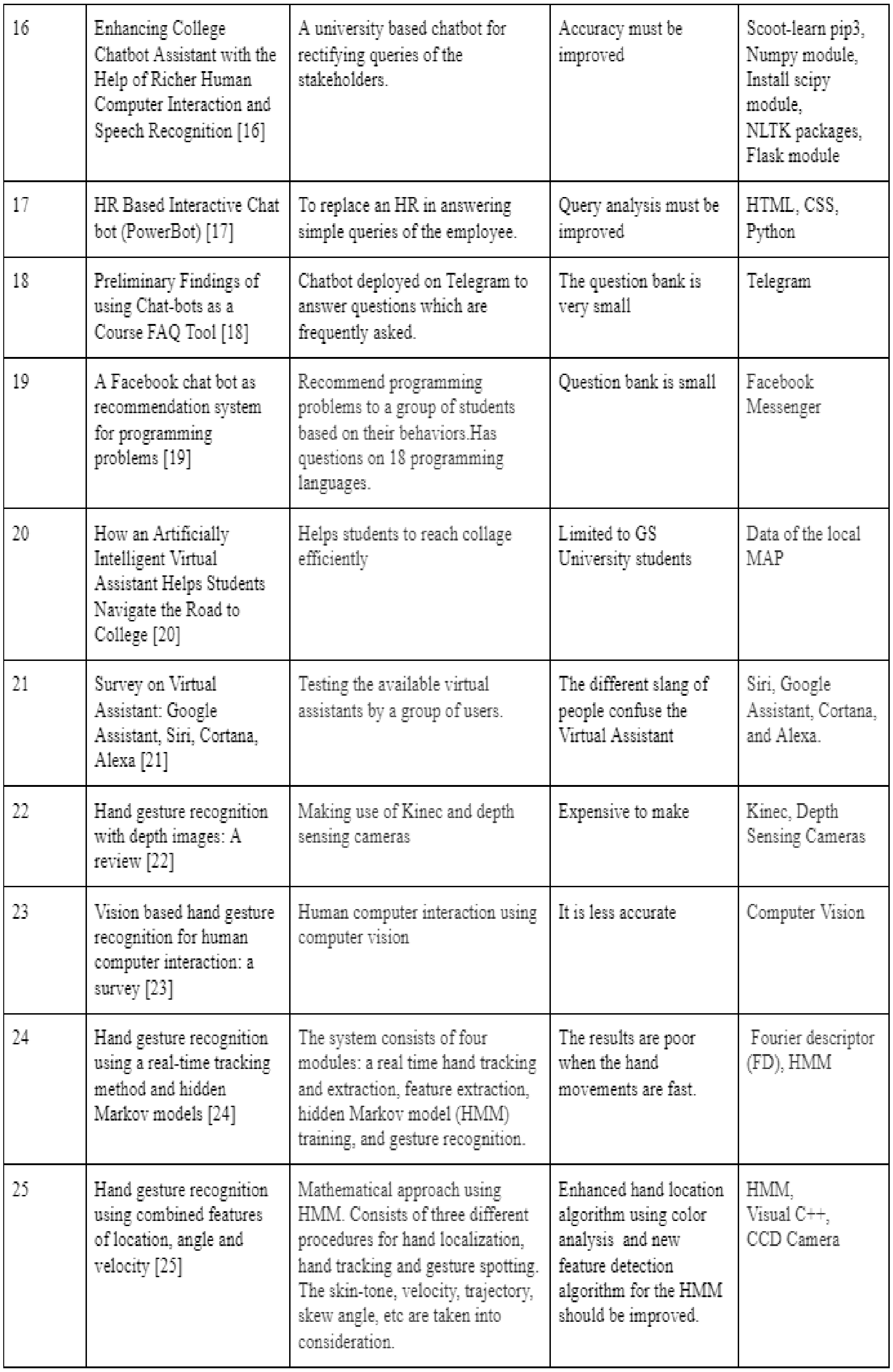
2. LITERATURE REVIEW



S.no	Title	Findings (Merits)	(Demerits)	Tools/ Concept
1	Adrisya Sahayak: A Bangla Virtual Assistant for Visually Impaired [1]	Developing a hands free virtual assistant for native Bangla speaking people with basic functions.	It supports only Bangla language and also the accuracy is not good.	Google Web Speech API, Text to speech API, Bangla vocabulary database
2	Implementation of Virtual Assistant with Sign Language using Deep Learning and TensorFlow [2]	An interface for recognition of sign languages of the deaf and converting them to speech.	It uses alexa for their searches and returns only text. It doubles the complexity of the virtual assistant.	Tensorflow, CNN, Python Audio modules
3	Sign Language Alphabet Recognition Using CNN [3]	Recognize American sign languages using Computer Vision.	The accuracy in spotting the hand is low.	MNIST classic dataset, Keras, CNN
4	Gesture recognition for sign language Video Stream Translation [4]	Sign language recognition framework based on attention mechanism		CNN LSTM Gradient weighted class activation
5	Thai Sign Language Recognition: an Application of Deep Neural Network [5]	Recognize Thai Sign Language using custom Dataset.	It supports only thai language and the dataset is not that good.	Mediapipe framework RNN, LSTM, BLSTM, GRU
6	Pravas Sarthi - A Convenience MultiLingual Virtual Assistant [6]	A Virtual assistant with multiple language support and GUI interface to guide passengers stations	It is a very simple interface.	Python SQLite DB
7	Efficient corpus design for wake-word detection [7]	Conducted an empirical study on design of corpora for wake-word detection.evaluated the performance of four neural network models, namely LSTM, CNN, CRNN, and DS-CNN, with eight different subsets of the corpus using EER, FAR, FRR and pAUC.	Focuses on a predefined wake-word. Applying the method to different wake-words is difficult.	LSTM, CNN, CRNN, and DS-CNN



8	Recipe for Creating a Highly Accurate Wake Word Engine [8]	Studied the previously less-studied augmentation technique of Echo audio transformation effect. Study on the impact of SpecAugment on WWS (Wake Word Engines)	Search space in arriving at the optimum combination of the augmentation techniques can grow exponentially	Echo audio transformation effect
9	Direct modeling of raw audio with DNNS for wake word detection [9]	Technique for training features directly from the single-channel speech waveform. computationally efficient DNN architecture	FIR filter is estimated solely from the training data so that the phones are well discriminated at each frame	LFBE(log-mel filter bank energy) , DNN(deep neural network)
10	Federated Learning for Keyword Spotting [10]	Showed that a revisited *Federated Averaging* algorithm with per-coordinate averaging based on Adam in place of standard global averaging allows the training to reach a target stopping criterion of 95% recall per 5 FAH within 100 communication rounds on our crowdsourced dataset	Frame labeling strategy used in this work relies on an aligner, which cannot be easily embedded.	*Federated Averaging* algorithm
11	My Eyes Are Up Here: Promoting Focus on Uncovered Regions in Masked Face Recognition [11]	Improving the accuracy of masked face detection.	Not as accurate as unmasked face detection.	traditional triplet loss, the mean squared error
12	Enhanced AlexNet with Super-Resolution for Low-Resolution Face Recognition [12]	An enhanced AlexNet with batch normalization and dropout regularization is then used for feature extraction.	Test accuracy is low.	AlexNet with Super-Resolution and Data Augmentation, k-Nearest Neighbors method
13	Morphological Preprocessing for Low-Resolution Face Recognition using Common Space [13]	Uses morphological preprocessing method for low-resolution face recognition, methods such as SR-net to increase image resolution, CNN for features extraction, and kNN for classification.	Works only with discrete images.	CNN, KNN, SR-net
14	Design of Embedded Intelligent Face Recognition Access Control System [14]	Introduces the working principle of the embedded intelligent face recognition access control system, the hardware and software of the system and the software flow.	Not widely used	VS2015 on Windows
15	HR-Chat bot: Designing and Building Effective Interview Chat-bots for Fake CV Detection [15]	Search through the CV and also use a face detection based system. Also has a chatbot.	No feature to assess the technical skills.	Naive Bayes model, List model, and Conditional Random Field model.



16	Enhancing College Chatbot Assistant with the Help of Richer Human Computer Interaction and Speech Recognition [16]	A university based chatbot for rectifying queries of the stakeholders.	Accuracy must be improved	Scoot-learn pip3, Numpy module, Install scipy module, NLTK packages, Flask module
17	HR Based Interactive Chat bot (PowerBot) [17]	To replace an HR in answering simple queries of the employee.	Query analysis must be improved	HTML, CSS, Python
18	Preliminary Findings of using Chat-bots as a Course FAQ Tool [18]	Chatbot deployed on Telegram to answer questions which are frequently asked.	The question bank is very small	Telegram
19	A Facebook chat bot as recommendation system for programming problems [19]	Recommend programming problems to a group of students based on their behaviors.Has questions on 18 programming languages.	Question bank is small	Facebook Messenger
20	How an Artificially Intelligent Virtual Assistant Helps Students Navigate the Road to College [20]	Helps students to reach collage efficiently	Limited to GS University students	Data of the local MAP
21	Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa [21]	Testing the available virtual assistants by a group of users.	The different slang of people confuse the Virtual Assistant	Siri, Google Assistant, Cortana, and Alexa.
22	Hand gesture recognition with depth images: A review [22]	Making use of Kinec and depth sensing cameras	Expensive to make	Kinec, Depth Sensing Cameras
23	Vision based hand gesture recognition for human computer interaction: a survey [23]	Human computer interaction using computer vision	It is less accurate	Computer Vision
24	Hand gesture recognition using a real-time tracking method and hidden Markov models [24]	The system consists of four modules: a real time hand tracking and extraction, feature extraction, hidden Markov model (HMM) training, and gesture recognition.	The results are poor when the hand movements are fast.	Fourier descriptor (FD), HMM
25	Hand gesture recognition using combined features of location, angle and velocity [25]	Mathematical approach using HMM. Consists of three different procedures for hand localization, hand tracking and gesture spotting. The skin-tone, velocity, trajectory, skew angle, etc are taken into consideration.	Enhanced hand location algorithm using color analysis and new feature detection algorithm for the HMM should be improved.	HMM, Visual C++, CCD Camera

25

3. PROPOSED WORK

Input - . After exploring different ways of implementation, the best one is used in the project. The system accepts input in text, audio and video input. It is up-to the user to select the mode. The UI of the assistant accepts the text which can be directly processed by the algorithm. If the user wants to input in audio format, then Speech Recognition API from google is used. The audio is converted to text and then it is processed. In the case of hand sign(Video) input the CNN is used to convert that to text.

Processing. The processing of the text is done using keywords from the text. The text goes through sequence of conditional statements and the most appropriate one is selected. For example, If there a keyword "play" , the assistant will look if there is something after that. If the second word is a name of a song, that song will be played.

Information. After exploring different python APIs the one which provided the best result for out test cases is selected. For the project the Wolfram-Alpha API is used as a first priority chat-bot as well as information system. If a satisfactory result is not obtained from Wolfram , the Wikipedia API will be used. Wikipedia has a big collection of information on various topics. And if the result is still not satisfactory, the assistant will open a fresh tab in the default web browser and show what google has.

Entertainment. The biggest music bank of the present world is YouTube. From the name of the song given by the user the best match will be found from YouTube and will be played. This can also play songs for different moods, different playlists, etc. The assistant can also play music or videos which are present in the local PC.

Wake Word Detection. Almost all assistants present today has a wake word to start a competition. For Alexa, it is "hey Alexa" and for google it is "hey Google". Neural Networks are employed for acoustic modeling these days. Since our main focus is on bringing up the deaf and the blind, the need for a wake hand gesture is also required. The users image will also be fed into the assistants face recognition part for recognizing the user and greeting them whenever they come.

Messaging. The assistant can currently text in WhatsApp and Email. This will be expanded to almost all the major messaging platforms. For this purpose the inbuilt python APIs are used . Python provides smtplib module, which defines an SMTP client session object that can be used to send mail to any Internet machine with an SMTP or ESMTP listener daemon.

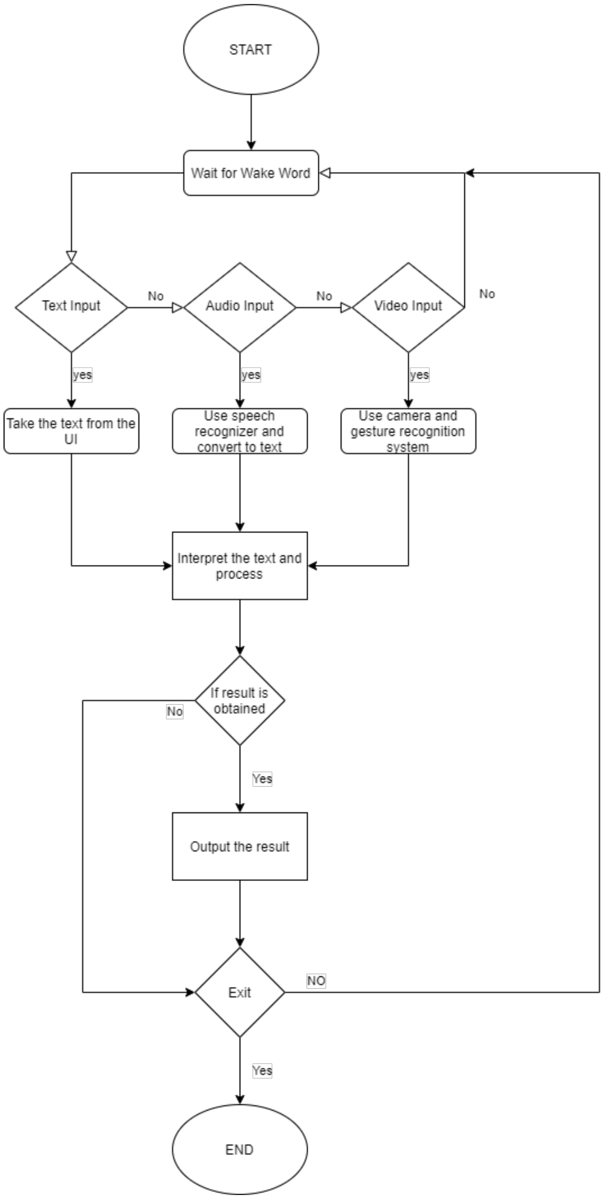

START
Wait for Wake Word
Text Input
No
Audio Input
No
Video Input
No
yes
yes
yes
Take the text from the UI
Use speech recognizer and convert to text
Use camera and gesture recognition system
Interpret the text and process
If result is obtained
No
Yes
Output the result
Exit
NO
Yes
END

Fig 1: Flow chart

4. References

[1] M. R. Sultan, M. M. Hoque, F. U. Heeya, I. Ahmed, M. R. Ferdouse and S. M. A. Mubin, "Adrisya Sahayak: A Bangla Virtual Assistant for Visually Im-

paired," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, pp. 597-602, doi: 10.1109/ICREST51555.2021.9331080.

[2] D. Someshwar, D. Bhanushali, V. Chaudhari and S. Nadkarni, "Implementation of Virtual Assistant with Sign Language using Deep Learning and TensorFlow," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 595-600, doi: 10.1109/ICIRCA48905.2020.9183179.

[3] M. Kumar, P. Gupta, R. K. Jha, A. Bhatia, K. Jha and B. K. Shah, "Sign Language Alphabet Recognition Using Convolution Neural Network," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1859-1865, doi: 10.1109/ICICCS51141.2021.9432296.

[4] B. Fei, H. Jiwei, J. Xuemei and L. Ping, "Gesture recognition for sign language Video Stream Translation," 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), 2020, pp. 1315-1319, doi: 10.1109/ICMCCE51767.2020.00288.

[5] A. Chaikaew, K. Somkuan and T. Yuyen, "Thai Sign Language Recognition: an Application of Deep Neural Network," 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, 2021, pp. 128-131, doi: 10.1109/ECTIDAMTNCON51128.2021.9425711.

[6] G. Bhatia, H. Tewani, A. Gunda, S. Kamat and A. Shankar, "Pravas Sarthi - A Convenience: MultiLingual Virtual Assistant," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-5, doi: 10.1109/ICCCNT45670.2019.8944428.

[7] D. Hossain and Y. Sato, "Efficient corpus design for wake-word detection," 2021 IEEE Spoken Language Technology Workshop (SLT), 2021, pp. 1094-1100, doi: 10.1109/SLT48900.2021.9383569.

[8] B. Ramanan, L. Drabeck, T. Woo, T. Cauble and A. Rana, "Recipe for Creating a Highly Accurate Wake Word Engine," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 4734-4740, doi: 10.1109/BigData50022.2020.9378193.

[9] K. Kumatani et al., "Direct modeling of raw audio with DNNS for wake

word detection," 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2017, pp. 252-257, doi: 10.1109/ASRU.2017.8268943.

[10] D. Leroy, A. Coucke, T. Lavril, T. Gisselbrecht and J. Dureau, "Federated Learning for Keyword Spotting," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 6341-6345, doi: 10.1109/ICASSP.2019.8683546.

[11] P. C. Neto et al., "My Eyes Are Up Here: Promoting Focus on Uncovered Regions in Masked Face Recognition," 2021 International Conference of the Biometrics Special Interest Group (BIOSIG), 2021, pp. 1-5, doi: 10.1109/BIOSIG52210.2021.9548320.

[12] J. C. Tan, K. M. Lim and C. P. Lee, "Enhanced AlexNet with Super-Resolution for Low-Resolution Face Recognition," 2021 9th International Conference on Information and Communication Technology (ICoICT), 2021, pp. 302-306, doi: 10.1109/ICoICT52021.2021.9527433.

[13] G. Marzani, N. Suciati and S. C. Hidayati, "Morphological Preprocessing for Low-Resolution Face Recognition using Common Space," 2021 International Conference on ICT for Smart Society (ICISS), 2021, pp. 1-5, doi: 10.1109/ICISS53185.2021.9533241.

[14] N. Yang, "Design of Embedded Intelligent Face Recognition Access Control System," 2021 International Wireless Communications and Mobile Computing (IWCMC), 2021, pp. 1189-1192, doi: 10.1109/IWCMC51323.2021.9498683.

[15] D. S. AbdElminaam et al., "HR-Chat bot: Designing and Building Effective Interview Chat-bots for Fake CV Detection," 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), 2021, pp. 403-408, doi: 10.1109/MIUCC52538.2021.9447638.

[16] S. Kumari, Z. Naikwadi, A. Akole and P. Darshankar, "Enhancing College Chat Bot Assistant with the Help of Richer Human Computer Interaction and Speech Recognition," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 427-433, doi: 10.1109/ICESC48915.2020.9155951.

[17] S. Tadvi, S. Rangari and A. Rohe, "HR Based Interactive Chat bot (PowerBot)," 2020 International Conference on Computer Science, Engineering

and Applications (ICCSEA), 2020, pp. 1-6, doi: 10.1109/ICCSEA49143.2020.9132917.

[18] S. I. Ch'ng, L. S. Yeong and X. Ang, "Preliminary Findings of using Chat-bots as a Course FAQ Tool," 2019 IEEE Conference on e-Learning, e-Management e- Services (IC3e), 2019, pp. 1-5, doi: 10.1109/IC3e47558.2019.8971786.

[19] A. Prisco et al., "A Facebook chat bot as recommendation system for programming problems," 2019 IEEE Frontiers in Education Conference (FIE), 2019, pp. 1-5, doi: 10.1109/FIE43999.2019.9028655.

[20]. Page LC, Gehlbach H. How an Artificially Intelligent Virtual Assistant Helps Students Navigate the Road to College. AERA Open. October 2017. doi:10.1177/2332858417749220

[21] Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa Advances in Signal Processing and Intelligent Recognition Systems, 2019, Volume 968 ISBN : 978-981-13-5757-2 Amrita S. Tulshan, Sudhir Namdeorao Dhage

[22] J. Suarez and R. R. Murphy, "Hand gesture recognition with depth images: A review," 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, 2012, pp. 411-417, doi: 10.1109/ROMAN.2012.6343787.

[23] Rautaray, S.S., Agrawal, A. Vision based hand gesture recognition for human computer interaction: a survey. Artif Intell Rev 43, 1–54 (2015). https://doi.org/10.1007/s10462-012-9356-9

[24] Feng-Sheng Chen, Chih-Ming Fu, Chung-Lin Huang, Hand gesture recognition using a real-time tracking method and hidden Markov models, Image and Vision Computing,Volume 21, Issue 8,2003,Pages 745-758,ISSN 0262-8856,

[25] Ho-Sub Yoon, Jung Soh, Younglae J. Bae, Hyun Seung Yang, Hand gesture recognition using combined features of location, angle and velocity, Pattern Recognition, Volume 34, Issue 7,2001, Pages 1491-1501,ISSN 0031-3203,