iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** V **Month of publication:** May 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.42390

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue V May 2022- Available at www.ijraset.com*

# Desktop Voice Assistant

Ujjwal Gupta[1], Utkarsh Jindal[2], Apurv Goel[3], Vaishali Malik[4]
[1, 2, 3, 4]*Department of Computer Science and Engineering, Meerut Institute of Engineering and Technology*

***Abstract:*** ***The primary goal of trending technology artificial intelligence (AI) is to realize natural human-machine dialogue. Various IT-based companies also utilized dialogue networks technology to create various types of Virtual Personal Assistants focused on their products and areas for expanding human-machine contact, such as Alexa, Cortana, Google's Assistant, Siri and so more. Just like the Microsoft voice assistant named 'Cortana', we designed our virtual assistant which performs basic tasks based on the instruction provided to it on the Windows platform using Python. Here, Python is used as a scripting language as it has a large library that is used to perform instructions. Using Python packages, a personalized virtual assistant recognizes and processes the user's voice.***

***Voice assistants are a fantastic advancement in the sector of Artificial Intelligence that can transform people's lives in a variety of ways. The voice-based assistant was initially given on cellphones and quickly gained popularity. It was widely acknowledged by all. Previously, voice assistants were largely found in smartphones and laptops, but they are now increasingly available in various home automation setups and smart speakers. Many technologies seem to become wiser in their very own way, allowing them to converse with humans in a simple language. Desktop voice assistants are programme that can identify people's speech and answer through an integrated speech system. This paper will outline how different voice assistants work, as well as their primary challenges and limitations. The way of developing a voice-based assistant without requiring cloud services is discussed in this paper, which would promote the future growth of such devices.***

***Keywords: Voice Assistant, Speech Recognition, Python, Smtplib, Automation.***

## I. INTRODUCTION

Everything in the twenty-first century is trending toward automation, whether it's your home or your transportation. Over the past couple of years, there has been an incredible development, or rather advancement, in new tech. You can, presume it or not, but you can engage with your gadget in current period. What does it mean to engage with a machine? Providing it some input is obvious, but if the input data is not in the traditional method of typing, but rather your own voice? What if you communicate to the computer, give it commands, and just want the system to engage with you as if it were your private assistant? What about if the system does more than simply display user the best outcome? What about if it also advises them on a good option? The revolutionary method of human system interchange is to easily accessible machine via voice commands. To accomplish this, we must be using a API which converts voice messages into text messages to understand the input. Many companies, including Google, Amazon, and Apple, are attempting to attain this in a more universal manner. Isn't it great how you can make reminders simply by stating "reminds me to...." or "set an alarm" with wake me up at...? Recognizing the significance of this, we decided to create a platform that can be installed anywhere in the neighborhood and can be asked to assist anybody with anything simply by chatting with it. Furthermore, you can link two similar devices via Wi-Fi and have them interact with one another in the future. These devices can be highly useful for day-to-day use and can assist you perform better by providing you with frequent alerts and updates. Why else would we require it? As our own voice is becoming a better input device than just a standard enter key. All Operating Systems offer a plethora of apps and services to users. The most well-known iPhone application is "SIRI," which enables people to communicate with their phones via voice commands and responds to voice instructions. Google has also created a similar programme, "Google Assistant", which is utilized on Android smartphones. However, that application relies heavily on Internet connections. However, the proposed system may operate with or without using Internet connectivity, taking input from users in the form of speech or text and processing it before returning the outcome in various formats such as action to be taken. Voice-controlled based home automation technologies could provide consumers with a much more comfortable living and make routine tasks easier. Voice control in energy conservation building is especially advantageous for those with disabilities, allowing them to live a previously unattainable lifestyle. Implementing voice activated systems might provide significant benefits, including aid with tasks at work.

A voice assistant seems to be a computerized program agent that can execute tasks or provide services for a person through using voice control technologies. Several virtual voice assistants currently available are Amazon's Alexa, Microsoft's Cortana, Apple's Siri, Samsung Bixby, Google Assistant, and many other.

©IJRASET: All Rights are Reserved 

A voice-based assistant is a computerized program that executes programs or services which the user assigns to it using various instructions. In software jargon, the software agent that is accessed through live chat is described as a 'chatbot', and it is a part of the Digital agent category. Voice-based assistants in the same category can understand and respond to social speech.

Voice controlled systems enhances the conveniences given by such gadgets and has already been included in a number of systems. As example, the aforesaid driver could manage his vehicle's GPS device without taking his hands off the steering wheel, but the harried secretary might simply tell his smartphone to dial the number whilst also working on an important file. However, more technically proficient individuals may decide for using such a system since they choose talking instead of just typing, but just because it is more enjoyable.

## II. LITERATURE REVIEW

Raja N., Bassam A., and others have written on the most important remark and speech. The analogue signal was used to communicate between humans and machines, which was then converted into a digital wave via voice. This technology has been widely used; it has a wide range of applications and allows computers to respond to human voices in a continuous and suitable manner. It also provides useful and valued features. SRS (Speech Recognition System) is on the rise and has a wide range of uses. The procedure's summary was discovered through research; it is indeed a good model [1].

Speech analysis is commonly conducted in tandem with pitched analyses, as indicated by L. R. Rabiner and B. S. Atal. Based on the signal dimensions, the study developed a pattern identification method for identifying whether such a given segment of a voice signal must be classified as audio signal, unvoiced speech, or silence. The technique's principal limitation is the requirement to run the program on a specified variety of dimensions and under particular recording conditions [2].

Speech is the most common means of communication among humans, according to C. Vimala and V. Radha. Humans would prefer to communicate using machines using speech because that is the most advanced method. As a result, automated speech recognition has gained a lot of traction. The most common speech recognition methods are DTW(Dynamic Time Warping) and HMM. MFCC(Mel - Frequency cepstral Coefficients) were used for speech feature extraction because they provide a group of distinctive dimensions of sound signal. MFCC has been shown to be more exact and realistic then rest characteristics mining techniques in voice recognition in previous research. The work was conducted in MATLAB, and the results show that the machine is sensitive enough to detect words with a high level of accuracy [3].

A. Waiel and T. Schultz highlighted that when speech technological solutions proliferate over the world, the inability to adapt to new target languages becomes a useful worry. As a result, the study focuses on the question of how to quickly and efficiently convert LVCSR systems. Inside the context of the Worldwide Phone venture, which observes LVCSR techniques in 15 different languages, the study needs to reassess acoustic concepts for a novel target vocabulary using verbal information from various origin languages, but only limited data from the target language identifying outcomes that use language-dependent and independent, and linguistic acoustic concepts are explained and debated. [4].

Language is a fundamental medium of communication, according to J. B. Allen, and speech would be its primary interface. Speech signals were translated into analogue and digital wave shapes that a machine could understand as part of the human-machine interaction. [10] A technology that is widely used and has wide range of applications. Speech technology enables robots to respond to human words in a systematic and suitable manner, providing useful and desired services. The study includes an overview of the voice recognition process, its basic model, and applications, as well as a description of reasonable research into many strategies used in speech recognition systems. SRS is improving every time and has limitless applications. [5]

M. Bapat, P. Bhattacharyya, and others described a semantic analyzer for lots of the Indian language through NLP applications. [11] They defined and evaluated the semantic analyzer for "Marathi Language" during one of their projects. They began by devising a "boos trappable" encrypting approach that works even during function f, which is the technique's particular decryption function. The study found that the Marathi has a high level of correctness, with consistent derivational standards when using Finite State Systems to demonstrate language in a comprehensive fashion. Because Marathi has challenging semantics, clustering of post places and the formation of FSA is among the most important aids [6].

A prototype ASR for Bengali digit was published by M. N. Huda, G. Muhammad, and their colleagues. Despite the fact that Bengali is one of the most widely used languages on the planet, the collective works contain some of the few compositions of Bengali ASR, mainly Bengali accented in Bangladeshi. The amount is acquired from Bangladeshi citizens during this study. For detection, hidden Markov model (HMM) based attributes and MFCC dependent characteristics are used. As a result of performing deprivation, dialectic variance occurs. Gender-based testing revealed that female pronounced digits had higher accuracy levels as compared to male pronounced digits [7].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Sean R Eddy used different Hidden Markov methods, which are a standard statistical modeling method for 'linear' challenges like as sequencing or time - series data, and have also been widely used in voice recognition demands for the past two decades. It is possible to relate technical, entirely probabilistic procedures to profiling and pierced structure layouts using the HMM framework. [12] The majority of the difficulties associated with traditional profiles obtained have been addressed by profiles method built on the Hidden Markov model. HMMs provide a continuous structure for combining structural and sequencing data, as well as a stable concept for slotting insertions and deletions. Various sequence configurations based on HMM are rapidly being refined. Homolog significantly faster on HMM has previously been sufficiently effective for HMM methods to satisfactorily connect to far more challenging protein reversed fold threading procedures. [8].

## III. METHODOLOGY

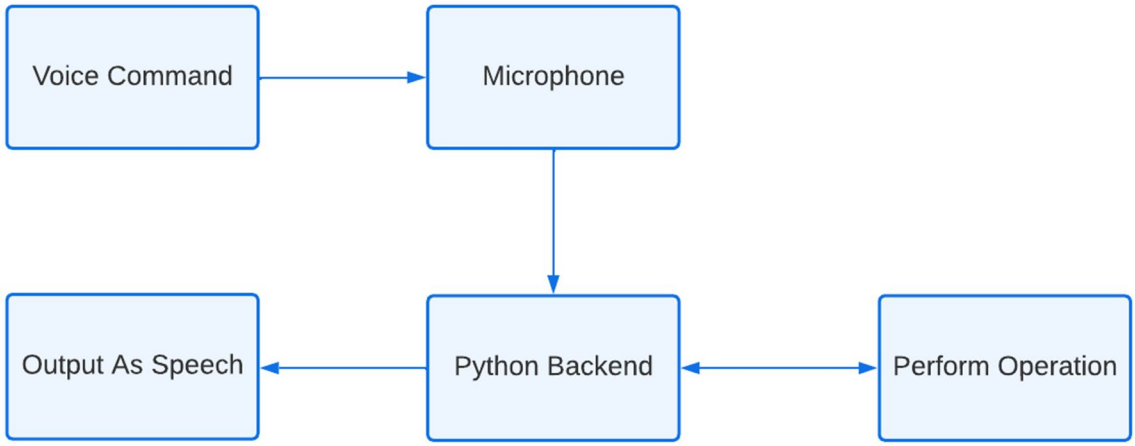



Figure 1: Basic Workflow of Model

### *A. Python Speech Recognition*

The device firstly converts speech input from the user into text using python module speech recognition algorithm. From voice input taken from users, we can obtain texts from specialized corpora arranged on the research center's computerized network server, that are briefly keep in the computer system before being transferred to python's module for recognizing speech. Then, the central processor accepts the similar text and feeds it.

### *B. Python Backend*

The backend of python parses the speech recognition module's response to examine that whether the speech or command outcome is either a System Calls, Send Mail, API Call, or Context Extraction. The data is then sent out towards the backend of python server to furnish the user with the relevant results.

### *C. API Call*

Term API stands for Application Programming Interface. The API is just a software-based interface which helps in establishing connection between two different systems at different locations. In plenty of other phrases, an API seems to be the messenger which sends your request towards the source location and afterwards returns the response to you.

### *D. Context Extraction*

Context Extraction (CE) is just the process of getting structured data from systems materials that are unstructured or semi-structured. The majority of the time, such activity entails using natural language processing to interpret human readable texts. TEST RESULTS for context extraction could be seen in current activities in visual report generation, such as content retrieval and automatic annotation from audio/images/video.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

*E. System Calls*

A system call seems to be the method using which a software programme asks for a service using the kernel of the computer's operating system where it is executing. Hardware-centric operations like, creation & execution of newly process, accessing a hard disc drive, the and communication between core kernel operations such as scheduling tasks are all examples of this. A process's interaction with the operating process is achieved by system calls.

*F. Python Text-To-Speech using Pyttsx3*

The extent of a systems to speak the provided text aloud is known as text to speech (TTS). Firstly, written file is transferred to a lexical representation, which is then subsequently converted to output waveform which can be used to create as sound file using a TTS Engine. 3rd-party authors offer TTS engine in a variety of dialects, languages, and specialist vocabulary.

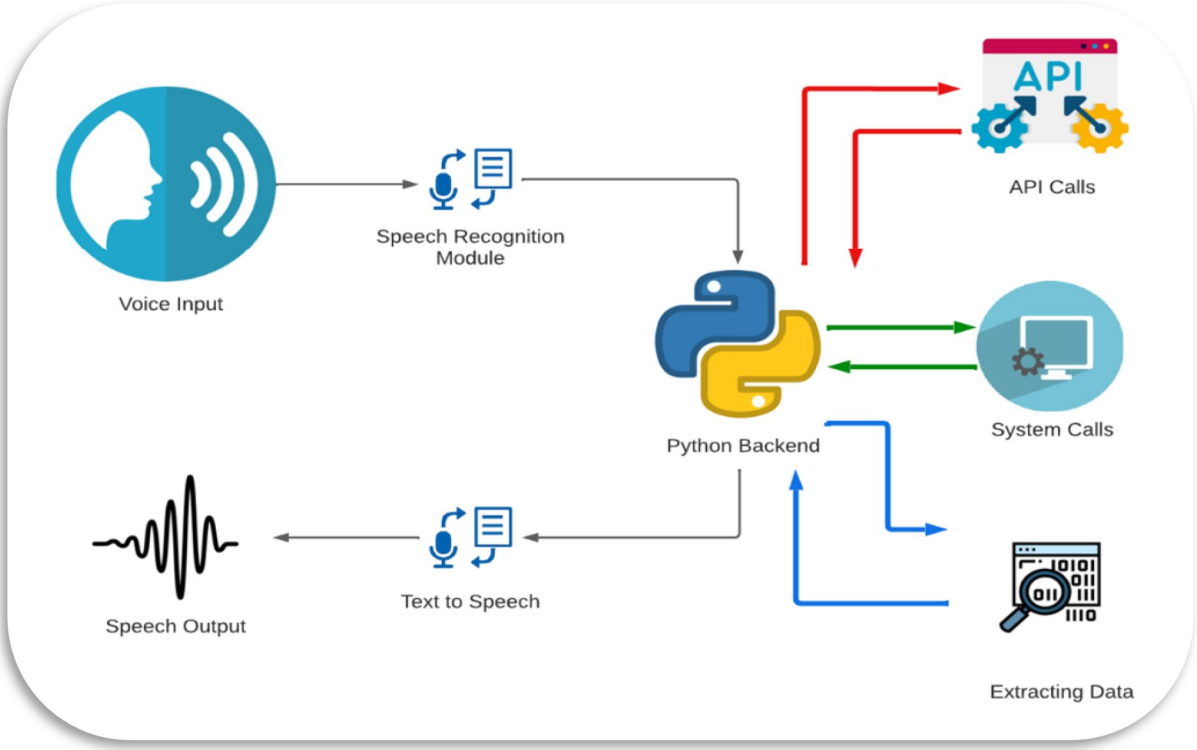



Figure 2: Detailed Workflow of Model

## IV. RESULT

A virtual assistant seems to be a quick and efficient aide. It's a piece of software that can decipher commands and carry out the tasks that the client has given it. NLP is used by virtual assistants to correlate user speech or text data with commands that may be executed. You may operate your devices like laptop or PCs using your own instructions with the use of virtual assistant. Because it is a quick process, it saves time. Because your virtual assistant works for you completes at set times, it will always be available with you and can swiftly adjust to changing demands. Your virtual assistant would be accessible to assist you and, if their demand permits, others such as relatives and coworkers.

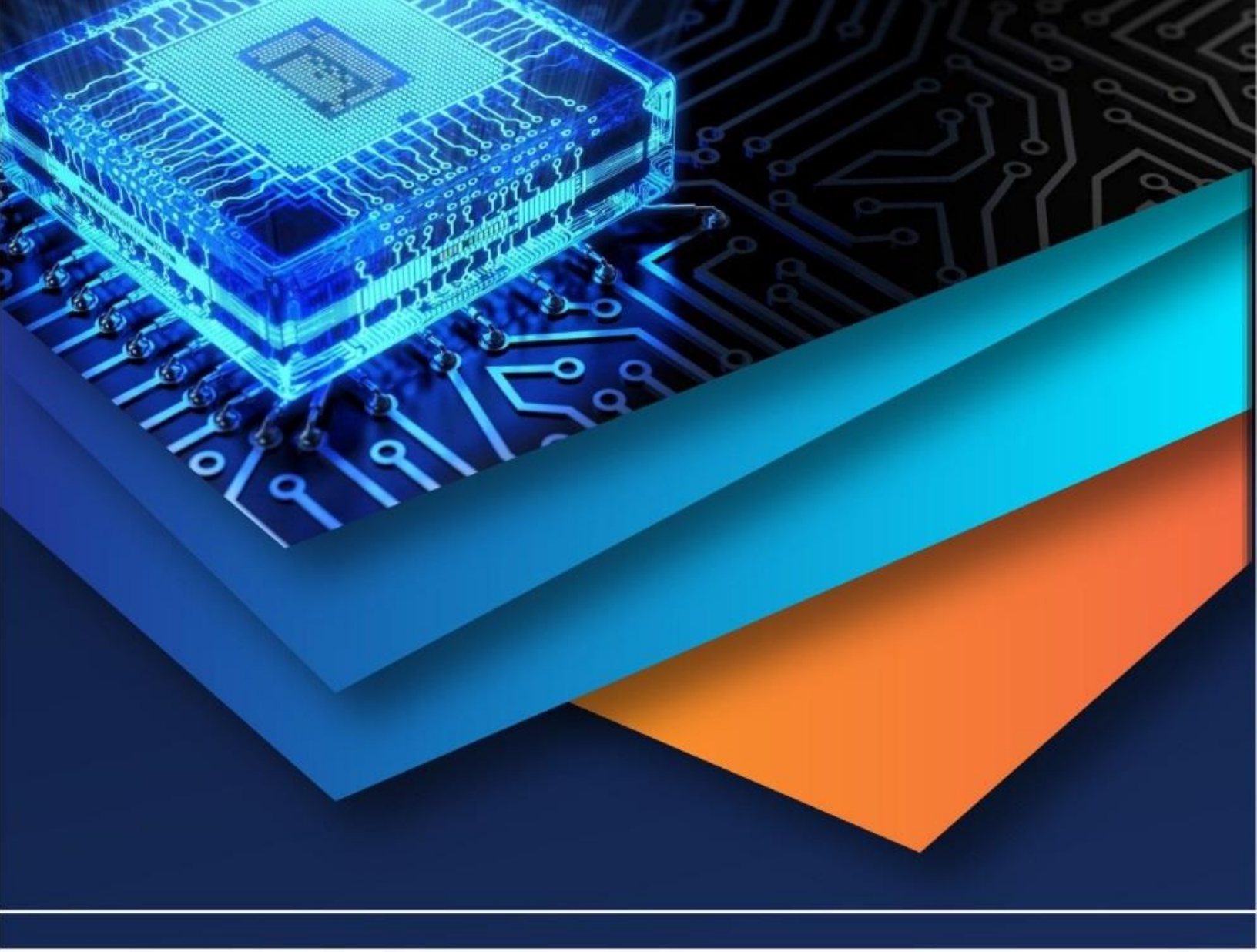
## V. CONCLUSION

We covered a Voice-operated Assistant written in Python in this paper. This assistant your basic tasks as a program that does basic activities like as providing weather updates, streaming music, searching on Wikipedia, and the opening of desktop apps, among others. The current system's capability is limited to solely working with applications. Artificial Intelligence will be incorporated into the system in future versions of this assistant, resulting in better recommendations with IoT to manage nearby gadgets, similar to what Amazon's Alexa does.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

## REFERENCES

[1] M. Bapat, H. Gune, and P. Bhattacharyya, “A paradigm-based finite state morphological analyzer for marathi,” in Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), pp. 26–34, 2010.
[2] B. S. Atal and L. R. Rabiner, “A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition,” Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 24, no. 3, pp. 201–212, 1976.
[3] V.Radha and C. Vimala, “A review on speech recognition challenges and approaches,” doaj. org, vol. 2, no. 1, pp. 1–7, 2012.
[4] T. Schultz and A. Waibel, “Language independent and language adaptive acoustic modeling for speech recognition”, Speech Communication, vol. 35, no. 1, pp. 31–51, 2001.
[5] J. B. Allen, “From lord rayleigh to shannon: How do humans decode speech,” in International Conference on Acoustics, Speech and Signal Processing, 2002.
[6] M. Bapat, H. Gune, and P. Bhattacharyya, “A paradigm-based finite state morphological analyzer for marathi,” in Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), pp. 26–34, 2010.
[7] G. Muhammad, Y. Alotaibi, M. N. Huda, et al., pronunciation variation for asr: A survey of the “Automatic speech recognition for bangla digits,” literature,” Speech Communication, vol. 29, no. in Computers and Information Technology, 2009. 2, pp. 225–246, 1999.
[8] S. R. Eddy, “Hidden Markov models,” Current opinion in structural biology, vol. 6, no. 3, pp. 361–365, 1996.
[9] Excellent style manual for science writers is “Speech recognition with flat direct models,” IEEE Journal of Selected Topics in Signal Processing, 2010.
[10] Srivastava S., Prakash S. (2020) Security Enhancement of IoT Based Smart Home Using Hybrid Technique. In: Bhattacharjee A., Borgohain S., Soni B., Verma G., Gao XZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science, vol 1241. Springer, Singapore. https://doi.org/10.1007/978-981-15-6318-8_44
[11] S. Srivastava and S. Prakash, "An Analysis of Various IoT Security Techniques: A Review," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 355- 362, doi: 10.1109/ICRITO48877.2020.9198027
[12] Saijshree Srivastava, Surya Vikram Singh, Rudrendra Bahadur Singh, Himanshu Kumar Shukla,” Digital Transformation of Healthcare: A blockchain study” International Journal of Innovative Science, Engineering & Technology, Vol. 8 Issue 5, May 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)