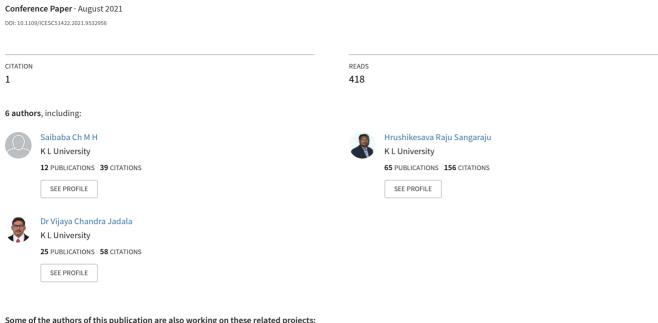
Intelligent Voice Assistant by Using OpenCV Approach



Some of the authors of this publication are also working on these related projects:



Mathematical Modeling for Defese System from Risks, Threats and Attacks View project

Intelligent Voice Assistant by Using OpenCV Approach

CH.M.H.Saibaba.

Assistant Professor,
Department of Computer Science & Engineering,
Koneru Lakshmaiah Education Foundation,
Greenfields, Vaddeswaram, Guntur – 522502.
Saibaba.ch77@kluniversity.in

S.Hrushikesava Raju
Department of Computer Science & Engineering,
Koneru Lakshmaiah Education Foundation,
Greenfields, Vaddeswaram, Guntur – 522502.
hkesavaraju@kluniversity.in

Vijaya Chandra Jadala
Associate Professor,
Department of Computer Science & Engineering,
Koneru Lakshmaiah Education Foundation,
Greenfields, Vaddeswaram, Guntur – 522502.
drvijayachandra@kluniversity.in

Abstract— The basic plan of our work is to integrate as several tasks as double and create it execute by our voice command. AN Intelligent Virtual Assistant (IVA) or 12 Intelligent Personal Assistant (IPA) could be a software system agent which will perform tasks or services for a private supported commands or queries. It is convenient, there are square measure some sectors wherever voice is that the solely approach of double communication, and a lot of typically, it permits to free-up each hands and vision doubtless for doing another activity in parallel or helps conjointly disabled individuals.

Generally, to try and do such quite tasks we would like a voice assistant to prefer and to get gadgets like Alexa. However identical practicality will be done by several the powerful packages like pywhatkit, Wikipedia, pyttsx3, pygame, speech recognition, OpenCV etc. [1] One of the relevant trends in artificial intelligence is that the technology of recognizing the natural language of a personalities. New insights during this topic will cause new means of natural human-machine interaction, during which the machine would find out how to know human's language, adjusting and interacting in it. One is every of such tools is voice assistant, which can be integrated into several alternative intelligent systems. Voice may be a ton of economical than writing on a keyboard. Here we tend to square measure desegregation varied options and place along as one practicable file. Hence, we will create several things among less quantity of your time with the simplest performance. The basic set up of our work is to integrate as many tasks as achievable and build it execute by our voice command.

Saiyed faiayaz waris,
Assistant Professor,
Department of Computer Science & Engineering,
Vignan's Foundation for Science Technology and
Research,

Guntur – 522213. Saiyed.cse@gmail.com

VSRK Sarma,

Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur – 522502. sharmavsrk@kluniversity.in

CHITTURI PRASAD,

Student,

Department of Computer Science & Engineering, Gandhi Institute of Technology and Management, Gandhi Nagar, Visakhapatnam, Andhra Pradesh 530045. chprasadklef@gmail.com

Keywords - Computer Vision; OpenCV; Pywhatkit, Wikipedia, Pyttsx3, Pygame, Speech Recognition.

I. INTRODUCTION

Intelligent Voice Assistants are commonly cloud based projects that involve web related gadgets as well as appliances to work. Three such applications are Siri on Apple devices, Cortana on Microsoft Devices and Google Assistant on Android devices. There are additionally devices committed to giving virtual help. The most renowned ones are available from Amazon, Google, and Microsoft. To make use of the Amazon Echo helper, called Alexa, clients get down on the wake word, "Alexa." A light on the device signs to the client it is prepared to get an order, which normally consist of basic language demands, for example, "what is the climate today," or "play popular music." Those solicitations are treated and put away in Amazon's cloud.

The advancements that power assistants require gigantic measures of information, which takes care of manmade brainpower (AI) stages, including AI, normal language preparing and discourse acknowledgment stages. [2] A major limitation of those merchandise is that speech recognition is performed on a server. Mobile network connections square measure often slow or intermittent, and generally non-existent. As the end client connects with a mental helper, the

AI programming utilizes modern calculations to gain from information include and turn out to be better at foreseeing the end client's requirements. [3] The work on creating and improving such personalized assistants has been going on for a long time. These systems are constantly improving and improving, go beyond personal computers and have already firmly established themselves in various mobile devices and gadgets.

In our task, we mostly use voice as correspondence implies, so the fundamentally the Speech acknowledgment application. Same sort of utilization is likewise evolved by the Google that is "Google Voice Search" which is utilized for in Android Telephones. Yet, this application generally works with Internet Associations.[4] There are square measure a range of industrial and ASCII text file systems such as AT&T Watson, Microsoft API Speech, Google Speech API, Amazon Alexa API, shade Recognizer, WUW, HTK and Dragon. A discourse synthesizer takes as info and produces a sound stream as yield. A discourse recognizer, then again, does the direct inverse. It takes a sound stream as information what is more, along these lines transforms it into text record. [5] Nowadays, most current AI researchers work instead on tractable "narrow AI" applications (such as diagnosing or automobile navigation). Consequently, the computerized signal cycles like Feature Extraction and Feature Coordinating are acquainted with address the voice signal. In this task, we straightforwardly utilize the discourse motor which employments highlight extraction method. How cool is it to fabricate your very own partners like Alexa or Siri? It is not exceptionally muddled and can be effectively accomplished in Python. Individual computerized aides are catching a great deal of consideration recently. Chatbots are regular in most business sites.

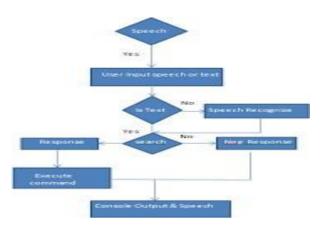


Fig 1-Standard Tracking Architecture

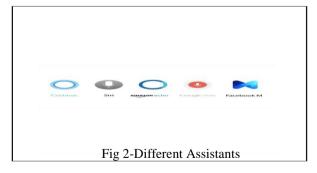
Google's content to-discourse bundle, gTTS changes over your sound inquiries to message. The reaction from the look-into work that you compose for bringing answer to the

inquiry is changed over to a sound expression by gTTS. This bundle interfaces with Google Translate's API. Playsound bundle is utilized to offer voice to the response. Playsound permits Python to play MP3 records. Internet browser bundle gives an undeniable level interface that permits showing webbased pages to clients. Selenium is another alternative for showing site pages. Be that as it may, for utilizing this you need to introduce and give the program explicit web driver. Wikipedia is utilized to get an assortment of data from the Wikipedia site.[6] Most existing chatbots accommodates dialog management modules to control the spoken language method and chatbot information bases to response to user input.[7] [9].

Referring vital analysis progress within the field of text to speech synthesis, the unit selection technique will provide most naturally sounding output speech. The set is chosen from a hard and fast of things, each having associate object sort a minimum of one taggable field is related to the article sort and has a corresponding worth. The set of objects is saved within the laptop computer memory.[8] Due to the amount of tasks that use voice commands, studies that arrange to live the effectiveness of those assistants or compare them tend to concentrate on a tiny low number of assistants and square measure targeted to a slim field of usage situations during which authors perform measurements by themselves of one item is retrieved from the set of gadgets, the item of the kind selected through the user and having a worth inside the taggable space choice that matches the taggable field fee obtained from the user the command is finished on the item.

1.1 VOICE ASSISTANT

The Voice Assistant Service supports the popularity of voice commands to electronic devices. Figure illustrates how the voice service works. To control a smart device, a user will speak a voice command. [10] All assistants square measure compared supported the same aspects and services. The Assistant then sends the sounds of that voice command to a foreign voice process cloud via its connected wireless local area network. The Speech Recognition library permits Python to get to sound from your framework's amplifier, interpret the sound, and save it.



Virtual Assistants are rapidly evolving to provide more power and value to customers. As general language

recognition and language management continue, so does the low helper's ability to understand and perform needs. In addition, as voice recognition development progresses, the use of remote services will further the business process processes.

The forthcoming Intelligent Assistants will be functioned with additional developed knowledgeable processing advances, which will realize and whole multistep requests and make more changeable assignments, like making a aircraft. Use pip introduce to introduce the libraries prior to bringing in them. The following are the portion of the key libraries utilized in this program: The Speech absolute plan comprises of these stages: 1) Assortment of information which is in discourse design. 2) Analyze the speech and alter it to message. 3) Putting away information & preparing it. 4) Dialogue age from the content yield that is handled.

Information Flow Sequence:

a. Instate gadget: Initialize the gadget by calling its name. b. Undertaking Manager: Conversion of Speech-to-Text and Text-to Speech is performed by task chief. c. Administration Manager: Analysis of orders and coordinating them with web administration and applications. d. Execute Command: After finding the counterpart for the given order, run the individual python content and give the yield.

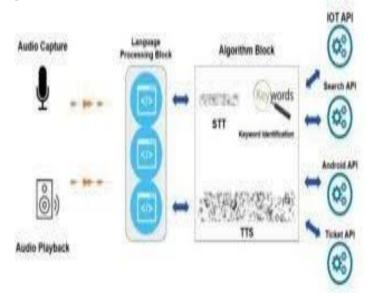


Fig3-Analysis of audio
II LITERATURE SURVEY

An AI individual collaborator is a piece of programming that comprehends verbal or composed orders and finishes task appointed by the customer. It is an illustration of feeble AI that is it can just execute and perform mission planned by the client. The python programming language, a content most usually utilized by the engineers can be utilized to construct your own AI right hand to perform task planned by the clients. Intelligent Voice Assistants regularly perform basic

locations for end clients, for instance, adding assignments to a agenda; giving data that would ordinarily be watched in an internet browser; or monitoring and scrutiny the situation with keen home devices, including lights, interior controllers, and cameras.

A. Speech Recognition

Clients additionally job the assistants to resolve on and get phone conclusions, make prompt communications, get bearings, hear bulletin and climate predictions, determine housings or coffee shop, check aircraft hesitations, listening melody, or muddle around. The carried-out voice associate can play out the accompanying errand it can open YouTube, Gmail, Google chrome and stack flood. [11] Foresee ebb and flow time, snap a picture, search Wikipedia to digest required information, anticipate climate in various urban areas, get top feature news from Times of India and can respond to computational and geological inquiries as well as since it doesn't have physical presence based mostly access management, the device will still settle for voice commands although no folks square measure around it. [12] It works for all the sounds reaching it at the pressure level (SPL) 60dB or higher. Speech Recognition has a long history with a few floods significant advancements. Discourse acknowledgment for transcription, search, what's more, voice orders has become a standard component on cell phones and wearable gadgets. Plan of a smaller enormous jargon discourse acknowledgment the framework that can run productively on cell phones, precisely what is more, with low inertness.

This is accomplished by utilizing a CTCbased LSTM acoustic model which predicts context-independent telephones and is packed to a 10th of its unique size utilizing a blend of SVD-based pressure what's more, quantization. Quantized profound neural organizations (DNNs) what's more, on-the-fly language model rescoring to accomplish continuously execution on present-day cell phones.[13]

Characterize a capacity take command for the AI aide to comprehend and to acknowledge human language. The mouthpiece catches the human discourse, and the recognizer perceives the discourse to give a reaction. The special case taking care of is utilized to deal with the exemption during the run time blunder and, the recognize google work utilizes google sound to perceive discourse.

Voice assistants have many attention-grabbing competences such as: A) Reply to queries requested by workers. B) Set regulators or frights. C) Play sports. D) Create phone call or send messages. E) Create goods. F) Give info regarding the climate. Emerging technologies like video game, increased truth and voice communication or redesigning the approach folks interact with the globe and remodeling

numerical practices. Voice management is that the succeeding progress of human-engine interaction, due to advances in cloud calculation, computing (AI) and also the Internet-of-Things (IoT). [14] In the previous ages, the serious use of smart applicances light-emitting diode to the looks of voice supporters such as Apple's concept Siri, Google's concept Assistant, Microsoft's concept Cortana and Amazon's concept Alexa. Voice assistants use technologies like voice recognition, speech synthesis, and NLP to supply facilities to the employers.



Fig 4-Different Uses

B. Voice Management

A voice boundary is crucial for IoT strategies that lack bit abilities (Metz, 2014). Moreover smartphones, voice supporters are currently combined in devices that are furnished with a electro-acoustic transducer and a speaker to interconnect with the users, known as good chatterers. Cloud environments area unit currently sanctionative voice assistants in scores of homes. [15] Voice assistants deem a cloud based design, meanwhile knowledge has got to be directed back and onwards to central knowledge hubs. a wise utterer is comparatively easy purposely, which suggests furthermost of the computation and computer science process happens within the cloud and not within the device itself.[16]

A few people have communicated security worries about Virtual Assistants, like Amazon Alexa and Google Home, because these menial assistants need a portion of distinct data and are repeatedly "tuning in" to react to speech instructions. Remote assistants at that fact hold speech associations and distinct data to progress the customer practice Cortana, for example, works best by employing data from a client's appliance, comprising communications and dissimilar exchanges, a client's associates, area information, exploration account, and data from other Microsoft organizations and abilities - unknown requests - that customers decide to associate with. Customers can choose not to symbol in and distribute this data with Cortana, and alteration consents to maintain particular data from actuality collected, however these events restrict the remote assistant's helpfulness. Virtual

Assistant providers likewise retain up protection arrangements, which describe how each association usages and proposals distinct statistics.[17]



Fig 5- Basic system architecture

C. People having some disabilities

Few trainings targeted workers with psychological feature infirmities or vision difficulties, to research however voice boundaries will support them in their daily life and the way simple is for them to use voice supporters. Baldauf et al. (2018), considered speech based informal boundaries for the cognitively weakened, i.e. folks along medical specialty ailments with minor damage in instrumental actions of regular existing together with reading and texting difficulties. [18] They also note that requests of informal boundaries and voice supporters for folks through psychological feature impairments area unit scarce. They trust that voice helpers mustn't be inactive however to inductee a language with the users in certain situations. [19] They additionally signifies that voice supporters ought to be an accompaniment and not a additional to private contact with different persons. Authors achieve that voice supporters will act as a schmoozer as well as facilitate to reduce the sensation of aloneness.

III THEORITICAL ANALYSIS

Despite the fact that this definition centers around the computerized type of remote helpers, the term Virtual Assistant, or virtual helper, is likewise usually used to portray contractors who telecommute doing authoritative errands normally performed by chief associates or secretaries. Virtual Assistants can likewise be diverged from another sort of customer confronting AI programming, called keen counselors. Savvy counselor programs are subject-situated, while are task-arranged.[20]

The individual has no access to the voice processing cloud, the sensible home ability adapter, the device cloud, and sensible devices. Victims square measure the homeowners of Alexa devices. The individual doesn't need any physical access to the victims' Alexa devices or to be physically gift nearby them. they will ought to compromise indoor acoustic devices to play discretional sounds at victims' home, however, do not need to record, overhear, or replay victims' spoken voice commands.

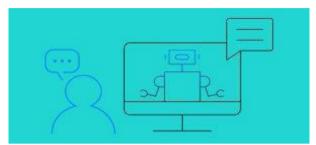


Fig 6- Human System Interaction

Note that our threat model mainly considers skulking attacks from the individual, we do not take into account invasive attack (where associate individual invades the victim's room) and crowd attacks (multiple adversaries attack at a similar time), as a result of too larger motions or too several adversaries might catch others attention.[21]

The accompanying orders assists with separating data from the wikipedia. The wikipedia.summary() work requires two contentions, the assertion given by the client, the number of sentences from wikipedia is should have been removed is put away in a variable outcome. The internet browser removes information from network. The open_new_tab work acknowledges URL as a boundary.[22]

An efficient power management system for the regulating the power flow in the lighting in the street must be capable of minimizing the power consumption. So the proposed method utilizes the ANN in the process of regulating the power flow to minimize the usage of the energy.[23]

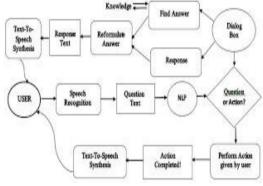
SDN platform assisted end-to-end delivery for social multimedia domain for detection of network anomaly is presented in this paper. Support Vector Machine (SVM) and Restricted Boltzmann machine (RBM) based ensemble approach is presented for detection of anomaly.[24]

The internet browser removes information from web. The open_new_tab work acknowledges URL as a boundary that should be gotten to. The current time is disconnected from datetime.now() work which shows the hour, moment and second and is put away in a variable name strTime. The

ec.capture() work is utilized to catch pictures from your camera. It acknowledges 3 boundaries:

- A) Camera list The initially associated webcam will be shown such as file 0 and the following webcam will be demonstrated such as record 1.
- B) Window name It can be a flexible or a string. In the event that you don't wish to view the window, treated as False.
- C) Save name A name can be provided to the picture and in the event that you would prefer not to save the picture, type as bogus.

Here we can utilize an outsider API known as Wolfram alpha API to answer computational and topographical queries. It is made conceivable by the Wolfram Language. The customer is an example (class) made for wolfram alpha. The res variable supplies the reaction given by the wolfram



alpha.

Fig7-Flow of Virtual Assistant

Virtual assistant is boon for everybody during this new era of twenty first century. it's made-up approach for a brand new technology wherever we are able to raise inquiries to machine and might act with IVAs as folks do with humans. This new technology attracted nearly whole world in many ways like sensible phones, laptops, computers etc. a number of the many VPs ar like Siri, Google Assistant, Cortana, and Alexa. Voice recognition, discourse understanding and human interaction ar the problems that aren't resolved nevertheless during this IVA's.

So, to resolve those problems a hundred users participated a survey for this analysis and shared their experiences. All users' task was to raise queries from the survey to all or any personal assistants and from their experiences this analysis paper came up with the particular results. in line with that results several services were lined by these assistants however still there are some enhancements needed in voice recognition, discourse understanding and

hand free interaction. once addressing these enhancements in IVAs will certainly augmented its use is that the main goal for this analysis paper.

IV EXPERIMENTAL INVESTIGATIONS AND RESULTS

In this work, we successfully built an Intelligent Voice Assistant that performs many functions that the current existing models cannot perform. The Intelligent Voice Assistant developed runs successfully and performs all the functions accurately and quickly making it a desirable model for future enhancement to a real-life software. The Intelligent Voice Assistant we created has functionalities that have never been introduced in any of the previously existing assistants available. Our model is not limited like the existing models and has a great scope for expanding its functionalities and adding new ones to the existing ones. There are many functionalities that are successfully executed by the Intelligent Voice Assistant built in the project. The functionalities have uniqueness when compared to the existing functionalities in the present current day models.

A) Search for a video in YouTube and play the video according to the highest rating.



B) This functionality listens to user request and converts in into a search request in Google.



Displays some technical jokes for the user when asked



A. Movement Detection feature:

This functionality detects the movement of the user and is also used as base for other functionalities of the Intelligent Voice Assistant. This functionality is crucial for implementation of other functionalities.



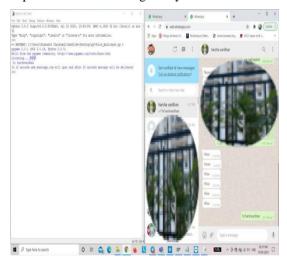
B. Automatic Scrolling in Social Media Websites:

This functionality module allows the user to scroll the social media websites by using a single color or face expression. As seen in the above result, we have set yellow color for the parameter to scroll while browsing through Facebook. When, the user displays yellow in front of camera, the site starts automatically scrolling reducing the use of mouse functionality and it also has a faster response speed when compared to the mouse scroll.



C. Sending a scheduled WhatsApp Message:

In this functionality, the Intelligent Voice Assistant automatically sends a Timed message through WhatsApp web. The user can set a certain time at which the message should be sent and to which person it should be sent. The Assistant will convert the recorded message to text and send it to the designated person at the time assigned by the user.



V CONCLUSION

We built an Intelligent Assistant with lesser limitations than the existing voice assistants. The project can be quite useful when incorporated into electronic gadgets. The Intelligent Voice Assistant performs various tasks asked by the user quickly and effectively making the project successful. This project can further be developed to make it useful to disabled people and allowing them to work with computers with ease. The voice recognition can be further developed to eliminate noise and then process the requests even in cases of high noised backgrounds.

Apart from this we also use various AI techniques to perform a variety of tasks by the assistant. It uses many python packages, Natural language Processing and image processing techniques to perform these tasks. The Voice Assistant surpasses the limitations of the existing models achieving a greater level of accuracy and ease of use along with increased functionality.

REFERENCES

- [1] Polyakov, E. V., Mazhanov, M. S., Rolich, A. Y., Voskov, L. S., Kachalova, M. V., & Polyakov, S. V. (2018, March). Investigation and development of the intelligent voice assistant for the Internet of Things using machine learning. In 2018 Moscow Workshop on Electronic and Networking Technologies (MWENT) (pp. 1-5). IEEE.
- [2] "Accurate and compact large vocabulary speech recognition on mobile devices," in INTERSPEECH. 2013, pp. 662–665, ISCA.
- [3] "AI Based Voice Assistant Using Python", International Journal of Emerging Technologies and Innovative Research (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.6, Issue 2, page no. pp506-509, February-2019, Available at: http://www.jetir.org/papers/JETIR1902381.pdf
- [4] G. Bohouta, V. Z. Këpuska, "Comparing Speech Recognition Systems (Microsoft API Google API And CMU Sphinx)", Int. Journal of Engineering Research and Application 2017, 2017
- [5] Artificial intelligence (AI), sometimes called machineintelligence. https://en.wikipedia.org/wiki/Artificial_intelligence
- [6] Huang, J., Zhou, M. and Yang, D., 2007, January. Extracting Chatbot Knowledge from Online Discussion Forums. In IJCAI(Vol. 7, pp. 423-428).
- [7] Mohasi, L. and Mashao, D., 2006. Text-to-Speech Technology in Human-Computer Interaction. In 5th Conference on Human Computer Interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI) (pp. 70-84).
- [8] Jiang, J., Hassan Awadallah, A., Jones, R., Ozertem, U., Zitouni, I., Gurunath Kulkarni, R., & Khan, O. Z. (2015, May). Automatic online evaluation of intelligent assistants. In Proceedings of the 24th International Conference on World Wide Web (pp. 506-516).
- [9] Fryer, L.K. and Carpenter, R., 2006. Bots as language learning tools. Language Learning & Technology.
- [10] Microsoft. Cortana. 21 May 2019. Available online: https://www.microsoft.com/windows/cortana/ (accessed on 28 October 2019).
- [11] Saibaba, CH. M. H., Uday Kumar, V., Praveenkumar, K., & Bhattacharyya, D, "A new privacy approach using graph-EMD". International Journal of Software Engineering and its Applications, Vol.10, No7, pp.11-16, August 2016.
- [12] Praveen Tumuluru, Burra Lakshmi Ramani, CH.M.H.Saibaba, B.Venkateswarlu, N.Ravindar, "OpenCV Algorithms for facial recognition", International Journal of Innovative Technology and Exploring Engineering, Vol.8, Issue.8, pp.927-933, June 2019.
- [13] Viranjitha Lakshmi Maturi, Nagarjuna Reddy Boya, Jaiakanth Polisetti, Sripujitha Adavi, CH.M.H.Saibaba, "Twitter Sentimental Analysis using Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering, Vol.8, Issue.6, pp.1592-1594, April 2019
- [14] K.Prasuna, K.V.S.N.Rama Rao, CH.M.H.Saibaba, "Application of Machine Learning Techniques in Predicting Breast Cancer A Survey", International Journal of Innovative Technology and Exploring Engineering, Vol.8, Issue.8, pp.826-832, June 2019.
- [15] Burra Lakshmi Ramani, Padmaja Poosapati, Praveen Tumuluru, CH.M.H.Saibaba, Mothukuri Radha, K.Prasuna, "Deep Learning and Fuzzy Rule-Based Hybrid Fusion Model for Data Classification", International Journal of Recent Technology and Engineering, Vol.8, Issue.2,pp3205-3213, July 2019.
- [16] Madhusudanan, J.; Geetha, S.; Venkatesan, V. Prasanna; Vignesh, U.; Iyappan, P., Hybrid Aspect of Context-Aware Middleware for Pervasive Smart Environment: A Review, MOBILE INFORMATION SYSTEMS, 2018, 10.1155/2018/6546501.
- [17] Maguluri, Lakshmana Phaneendra; Srinivasarao, Tumma; Ragupathy, R.; Syamala, Maganti; Nalini, "N. J., Efficient Smart Emergency Response System for Fire Hazards using IoT", International Journal of Advanced Computer Science and Applications, January, 2018.

- [18] Mekala, MS; Viswanathan, P, Sensor Stipulation with THAM Index for Smart Agriculture DecisionMaking IoT System, April, 2020.
- [19] Kumar, KG; Avinash, BS; Rahimi-Gorji, M; Majdoubi, J, Photocatalytic activity and smartness of TiO2 nanotube arrays for room temperature acetone sensing, February, 2020.
- [20] Sai, T. V. R., Haaris, S., & Sridevi, S, Website evaluation using opinion mining, International Journal of Engineering and Technology(UAE), 7(2), 51-53, 2018.
- [21] Sailaja, V. N., & Pavani, R., A study on understanding the relationship between personal attributes and investor perception towards ETF funds investment, Journal of Advanced Research in Dynamical and Control Systems, 10(8 Special Issue), 204-209, 2018.
- [22] Satish Babu, J., Ravi Kumar, T., & Bano, S, Optimizing webpage relevancy using page ranking and content based ranking, International Journal of Engineering and Technology(UAE), 7(2.7 Special Issue 7), 1025-1029, 2018
- [23] Smys, S., Abul Basar, and Haoxiang Wang. "Artificial Neural Network Based Power Management for Smart Street Lighting Systems." Journal of Artificial Intelligence 2, No. 01 (2020): 42-52
- [24] Chen, Joy Iong Zong, and S. Smys. "Social Multimedia Security and Suspicious Activity Detection in SDN using Hybrid Deep Learning Technique." Journal of Information Technology 2, no. 02 (2020): 108-115