Turkish Online Journal of Qualitative Inquiry (TOJQI)

Volume 12, Issue 3, July 2021:2108- 2121

Research Article

A Smart Home Automation with Personalised Voice Assistant Using IOT

Akurati Varun[1], Emani Akhil Reddy[2], K.M. Uma Maheshwari[3*]

Abstract

We are entering an advanced tradition where innovation is developing day - today. We pay special attention to greater thoughts that make life basic, simple, and security joins it. Distributed computing and IoT are electronic preparing where the information, application, and framework are given to PCs and different gadgets on-request over the organization. It is where the specialist co-op and the purchaser both can get the job done their business objective with the greatest benefits. Voice mechanization programming transforms discourse into activity orders or errands. This work points to one of the comparable applications to give subtleties on the most proficient method to manufacture an IoT Enabled voice mechanization utilizing minimal effort ware equipment and OpenCV programming. Voice mechanization innovation isn't great, in any case, and accompanies a couple of inconveniences like Lack of Accuracy and Misinterpretation of discourse acknowledgment of various accents and so forth The created framework can be coordinated as a solitary convenient unit and permits one to remotely control lights, fans, forced air systems, TVs, surveillance cameras, electronic entryways, PC frameworks, general media hardware's and so forth and turn ON or OFF any apparatus that is connected to a divider source, get the status of various sensors and take choice as needs be. The framework is convenient and built in a manner that is anything but difficult to introduce, design, run, and keep up. Adding highlights Like face acknowledgment and extra usefulness to administrators could upgrade the essential model in parts of security and protection. This framework diminishes human mediation, spares time, advances asset use. The assessment like example acknowledgment in the sound or discourse, assists with finding various clients from whom it got the orders and triggers the errands dependent on the examples in the orders.

Keywords: *Voice recognition, voice automation, arduino, IoT, artificial intelligence, machine learning, PIP (perceptual linear prediction), voice assistant.*

---

[1] UG Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India.

[2] UG Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India.

[3] Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India, umamahek@srmist.edu.in

Introduction

Voice recognition which started in the early 20th century for military purposes, in recognizing a few letters and characters is now spreading its branches into various fields like automation industries, authentication factors, etc. As the advancements in technologies are taking new turns that are helping people to achieve their targets effortlessly. Nowadays, this technology is made to reach every individual with an internet connection or a smartphone. Voice automation is one of the greatest enhancements that is ever going to be made in machine learning and artificial intelligence that lead to growth in house automation industries and building voice assistant. A portion of these home mechanization frameworks focus on those looking for extravagance and complex home robotization stages; others focus on those with exceptional requirements like the older and the impaired. As usage of these technologies not only has many benefits but also brings many security issues back to the home. One can easily get access to the previous commands or operations performed by the user. Factors like having the module accessible by everyone could be one of the reasons. One solution would be making the voice automation or voice assistant more personalized by maintaining the hierarchy in access to the voice assistant. In this survey, we will learn to build a personalized voice assistant, which will execute the commands spoken by the user after the recognition of the user using the voice recognition. Initially, we have to train the model upon the user's voices, this accomplished using many machine learning algorithms like logistic regression, Hidden Markov Model, Dynamic Time Warping. In this paper, we will be using ANN for voice recognition. In the case of speech recognition, deep learning algorithms like deep neural network, MFCC (Mel Frequency Cepstral coefficients) are used, in this survey paper, we will be using a PLP (Perceptual Linear Prediction) for speech recognition. In this survey paper, we will learn to build a voice recognition project that recognition speech and performs the different commands of the user that might have a similar meaning let's 'lighten up the room' on the 'lights in the room'. This paper helps in building the voice assistant which can dynamically help in home automation. This paper helps in enhancing the features of the voice assistant by integrating surveillance modal using Convolutional neural network (CNN).

Architecture Diagram of the Proposed System

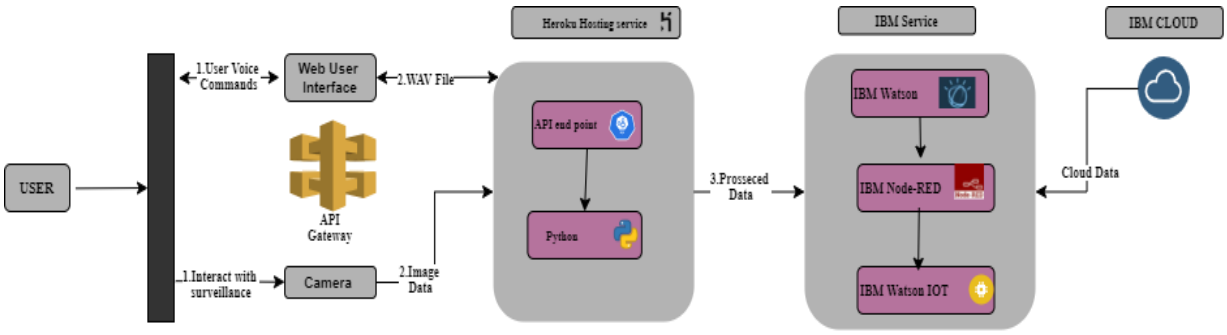


Figure 1. Architecture Diagram

The Architecture of the project is implemented in Three layers:

a. Web User Interface

This would be the top most layer, the Application Layer, where the end user interacts with a web application through mobile or on any device that has an internet connection and gives his/her voice commands to be executed.

The application layer will record the user voice commands and send a wav file to the next lower layer.

b. Python Backend

This would be the middle layer, the Business Layer, where the python backend code that get the wav file and process the file for user identification or processing the user commands like Lighten up the room and send a response text file to next lower layer.

c. IBM Service

This is the lowest layer, the Data Access Layer. This layer access the data from the cloud that is sent by the business layer and perform the cloud computation and send the response back to the business layer or to the user.

Related Works

The tasks which seems to be impossible around 19ths are now a piece of cake that are achievable in moment of eye blink. The evolutions in this technology era has turned many complex problems into a cake walk. Most of these problems are being solved using technological terms like Mobile application machine learning and artificial intelligence etc.

These software also turned handy in doing the tasks even without interaction of human beings. In a study, mobile application using the blue tooth in controlling the household electric devices like light fan and tv etcs[1].Making the modal more personalized, identification of the speaker even in the cases of speaker changes his/her languages is published in [2]. Usage of Google Api in speech recognition and machine learning algorithms in predicting the gestures improved the potential of the home automation industries [3]. Working on the face recognition and criminal detection will be expanding the potential usages of the voice assistant in homes[4]. Automation which involves in recognition of command based on the keywords in the speech makes the voice assistant more efficient and user friendly [5]. Cloud computation and services provided by the cloud helps in less process time for the modal at the device and helps in providing a vast number of features[6]. Identification of the face and localizing them with unsupervised learning algorithms can be seen in [7]. Wireless microcontrollers and software applications, that help in programming these micro controllers help in getting the weather condition whose data is further used in home automations[8]. Wake up words help in deciding the starting of speech and long silence can help is detecting the end of the speech[9]. Working on the voice assistant could be lot of challenging work, looking for the best architecture in building the modal is helps in solving those challenges[10].

Literature Survey

Survey Paper	Author	Algorithm	Merit	Demerit
1. Low-Cost Android App Based Voice Operated Room	1. Prasanmit Nath 2. Umesh Chandra Pati	PLP(Perceptual Linear Prediction)	1.Control home appliances like fan, bulb.	1.Only used for a certain range. 2. Cannot get

Automation System			2.Data transfers in fast	information from devices
2.Speaker Identification: Variations of a Human voice	1. Saritha Kinkiri 2. Simeon Keates	Mel Frequency Cepstral coefficients (MFCC) Dynamic Time Warping (DTW)	1.Achieved good accuracy in speech recognition 2.It works on extracting the features in the speech.	1.The main problem of, MFCC is that coefficients change when a speaker changes her/his language 2.problem is when several speakers increase
3.IOT BASED HOME AUTOMATION BY USING PERSONAL ASSISTANT	1. V.Chayapathy 2. B Sharath 3. G S Anitha	Deep Neural Networks Viterbi search WFST framework.	1.The focus of this work is to help users to operate home appliances with speech or gesture commands which give the user a better sense of control. 2.Less cost in building	1.Its takes a huge time in executing a command. 2.There is no authentication so anyone could access the modal and execute commands
4.CIS: An Automated Criminal Identification System	1. Kavushica Rasanayagam 2. W.A.D.D Tharuka 3. S.D.D.C Kumarasiri 4. Samanthi E.R. Siriwardana 5. Pradeepa Samarasinghe	Keras CNN TensorFlow Caffe deep learning framework	1.High accuracy in prediction 2.Takes less time to predict the person 3.Helps in predicting the emotion of the person on facial features.	1.Requires huge about of dataset to train upon. 2.The present system is only allowed to operate by system admin, but cannot send the information to the mobile of admin who might be in a remote location.
5.Investigation and Development of the Intelligent Voice Assistant for the Internet of Things Using Machine Learning	1. E. V. Polyakov 2. A. Y. Rolich 3. M. V. Kachalova 4. S. V. Polyakov	1. naive Bayesian classifiers, 2. decision trees	g1.synonyms and different variants of pronunciation of keywords, according to which the system will be able to build forecasts. 2.Less cost in building	1.Its takes a huge time in executing a command.
6.Internet of Things using Node-Red and Alexa	1. Anoja Rajalakshmi 2. Hamid Shahnasser	1. natural language processing	1.Usage of AWS services. 2. Node-Red provides ease for connections by just drawing wires or links and adding parameters to them.	1.Others can also access echo conversations.
7.An Unsupervised Learning Network for Face Identification and subsequent Localization	1. Goutam Sarker	1. Minimal Learning 2. Malsburg Learning 3. Leaky Learning	1.Performing task using different algorithms help in finding the best unsupervised machine learning 2.	1.Considering weighted average of different algorithms predicted makes the modal more efficient. 2.accommodate images has limit.
8.System of Wireless Temperature and Humidity Monitoring Based on Arduino Uno platform	1. Yanping Wang 2. Zongtao Chi	1. ATMEGA328 microcontroller is a high-performance single chip microcomputer	1.Simple to write code 2.Cost efficiency. 3. strong antiinterference and strong portability	1.ACR architecture is just for beginners not very useful for Bigger complex embedded applications.
9.One Solution for Voice Enabled Smart Home Automation System	1. Una RadosavacUna Radosavac 2. Milica Matić 3. Marija Antić	1. Google Speech API	1.Work on different languages. 2. The average time to get the response from the cloud recognition service is 1.72 seconds	1.Different Command parser are needed to be downloaded for XML configuration file 2.Errors can be caused by the MQTT message handling, as well as in the cases when the user has not correctly specified the name of device, or the desired action
10.Challenges of Integrating Smart Home Automation with Cloud-Based Voice Recognition Systems	1. Una Radosavac 2. Milan Vidaković 3. Milica Matić 4. Igor Stefanović	1. VRS cloud services	1.Discussed possible architectures of the HA system, which is using cloud-based VRS to provide voice command interface to the end-users. 2.The cost of bringing a voice command interface to the end-users can be lowered by using third party	1.The drawback of this architecture is the fact that command patterns have to be predefined with a great level of detail, and even device naming rules need to be provided. 2.this architecture requires implementing a more complex

			devices that can communicate with the VRS.	user authentication module within the HA cloud.

Features of the Proposed System

The project provides its services to the user in accomplishing his day to day task with at most efficiency and security by implementing the features like.

i. Smart Home Automation Using Voice Commands

The feature that gives the end user the personalised experience of a voice assistant built using the cloud services of IBM like node-red, watson. The module authenticates the user when he/she speaks the wake up word to the module like 'Hi Jarvis'. Ater the user is authenticated he/she can only perform the task or actions that they are authorized to do. This feature could even help in implementation of the voice automations in industry level,so that employers of a particular level can perform their task with huge efficiency and in less time.

ii. Smart Surveillance

The feature that gives the end user to have an additional security feature to his/her home. The module basically uses a camera which is attached at the front door to capture the images of intruders in case of any breaking. This feature could also help the module user to give the access to a known person to enter the house in cases where he/she cannot be present to welcome them.

Workıng Prıcıples of the Proposed System

i. Features of Smart Home Automation Using Voice Commands

a. Data Extraction

The voice utterance samples of the two people and collected speaking commands like Hi jarvis, Hello jarvis, ola jarvis in the wav (Wave form Audio File) format.WAV file helps to store the audio files in the digit format that can be used in feature extractions and module preparation for speaker identifications.

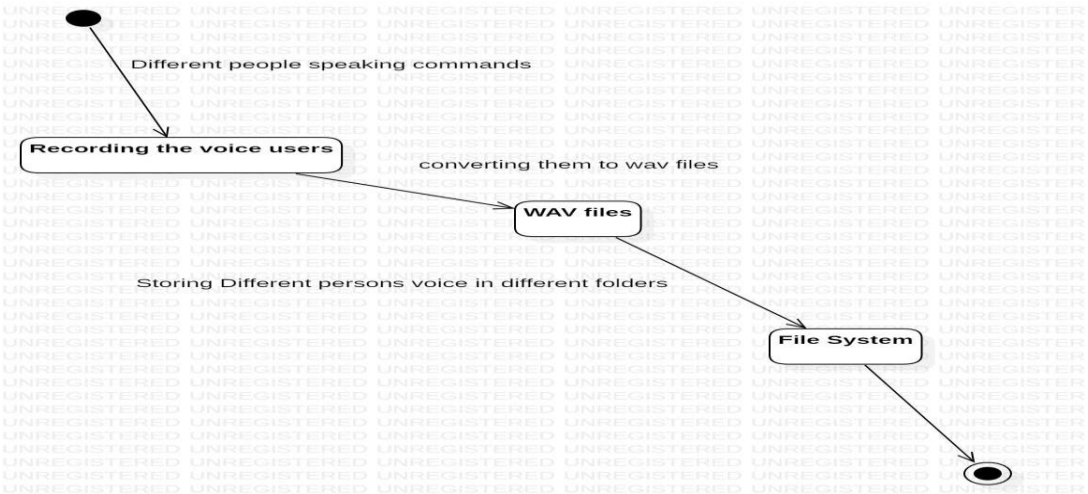


Figure 2. Data Extraction

b. Data Preparation Training

The audio files are preprocessed starting with silencing the starting and ending by considering the low amplitude frames as noise using the librosa python library. Then the features are extracted from the audio files using the mfcc algorithm and saving the features into dataframes using the Pandas library. The data frames are like table format where the features are stored and these data frames are given as input to deep learning algorithms like ANN.

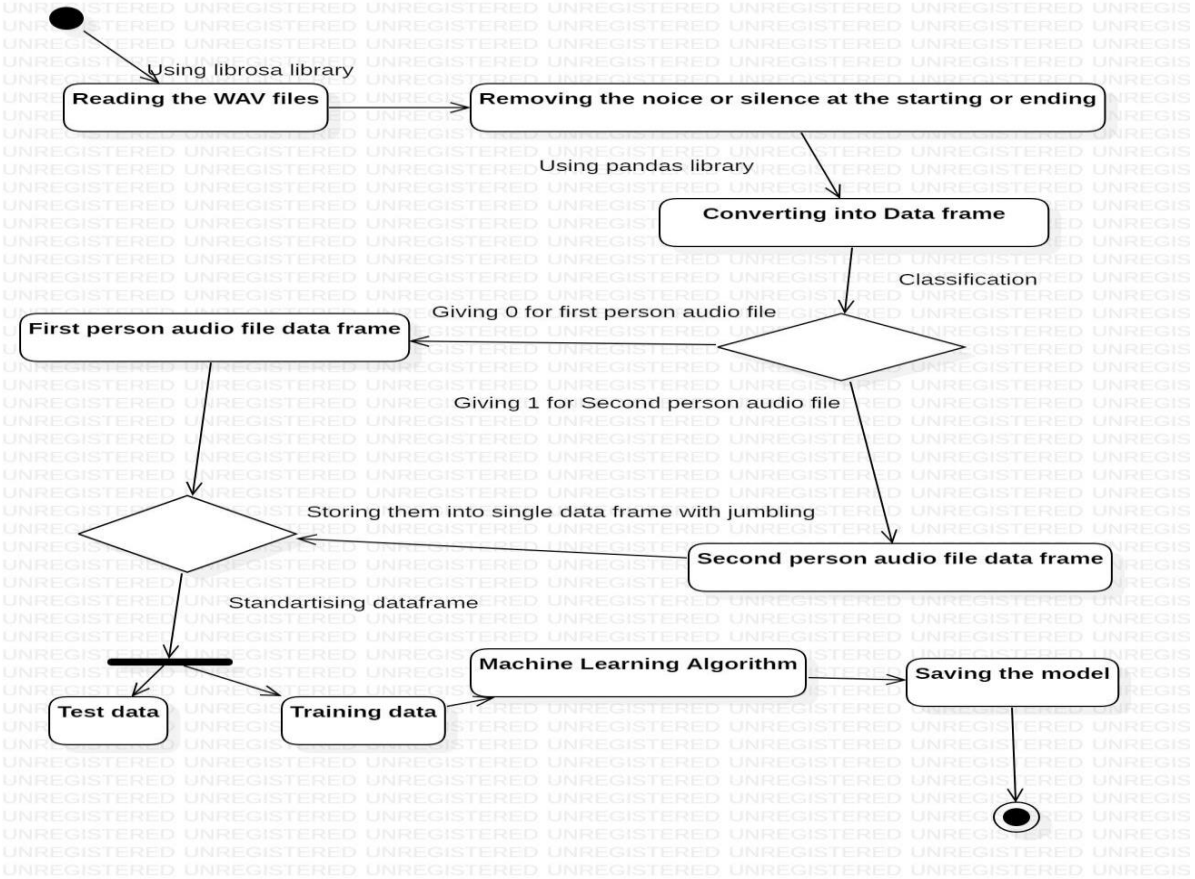


Figure 3. Data Preparation and Training

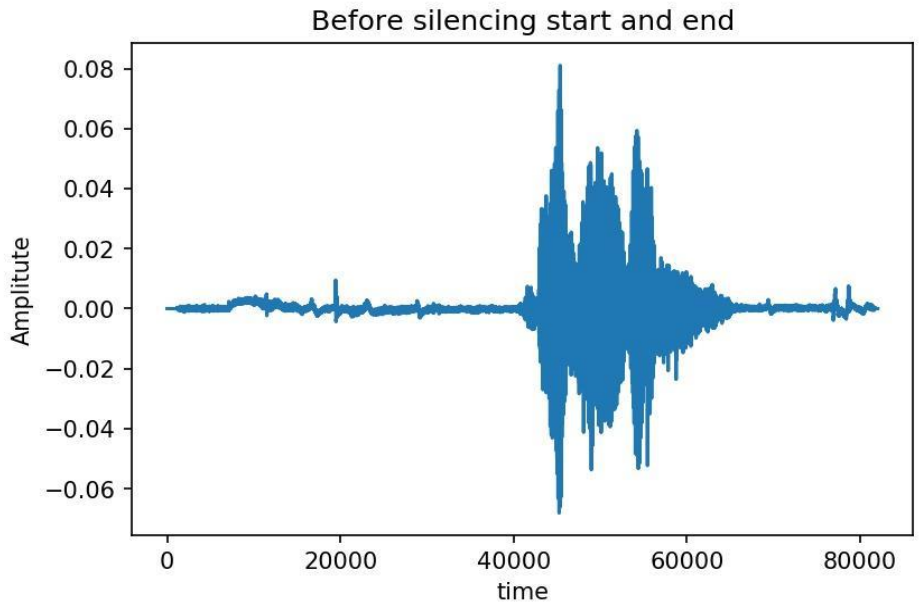


Figure 4.1. Before Silencing

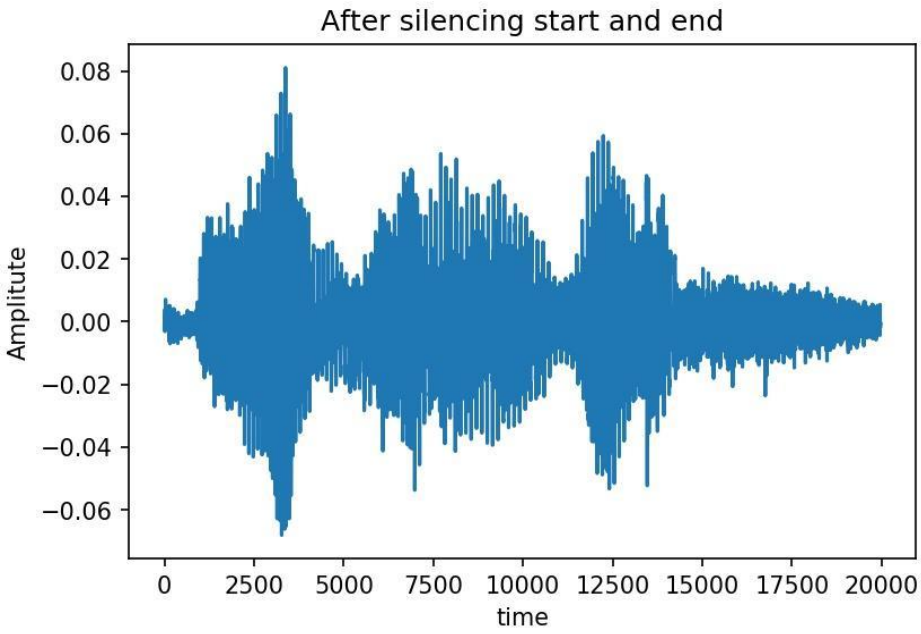


Figure 4.2. After Silencing

c. Data Prediction

When the user speaks the wake up word, it is recorded as wav format and given to a trained speaker identification model after processing the audio file.

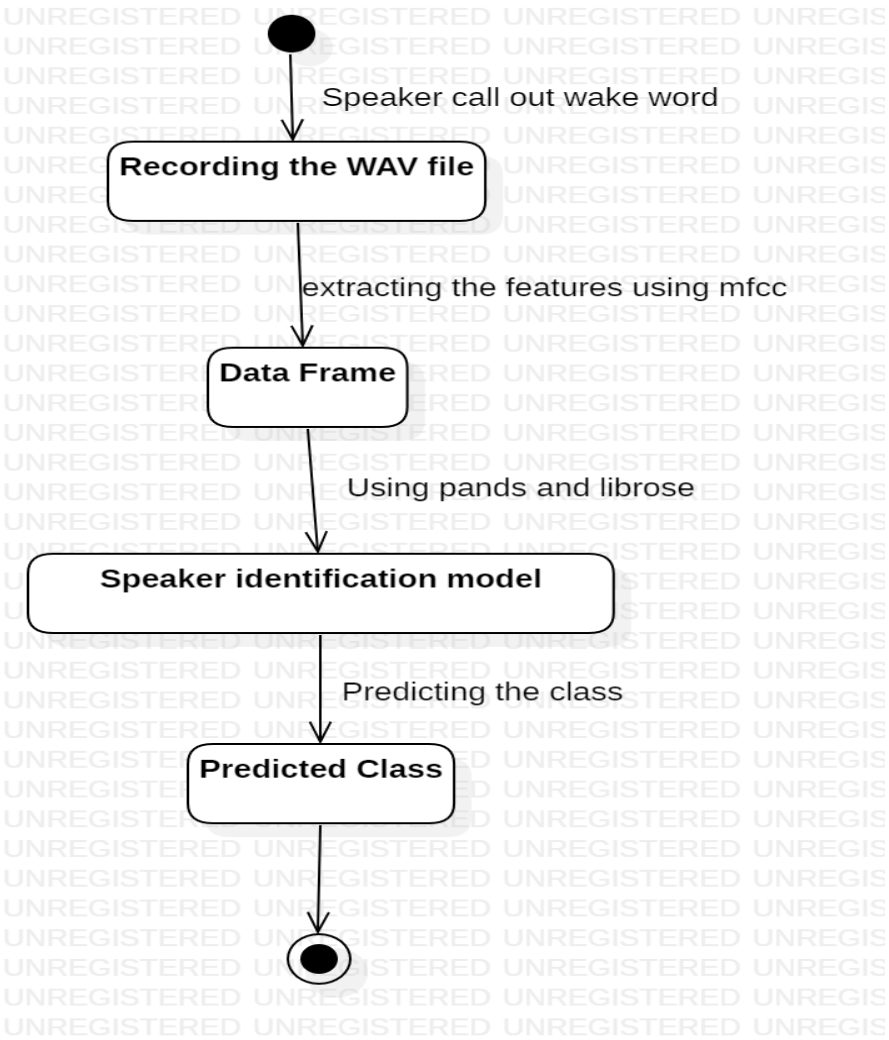


Figure 5. Data Prediction

ii. The Working Principles of Smart Surveillance

a. Data Extraction for Face Recognition

A sample video recording of 20 sec of the three people is collected with a person rotating his head in different directions and exposing various face gestures like smile anger.

The frame rate of the video file is set to 23fps. By setting the frame rate we can reduce the redundant images from the data set, so that the model overfitting on the training data can be reduced.

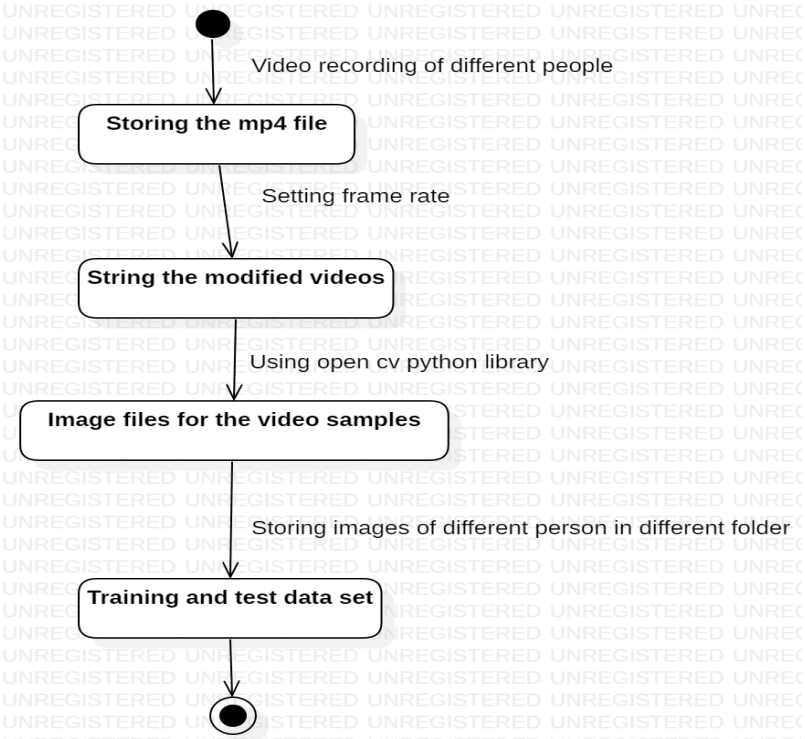
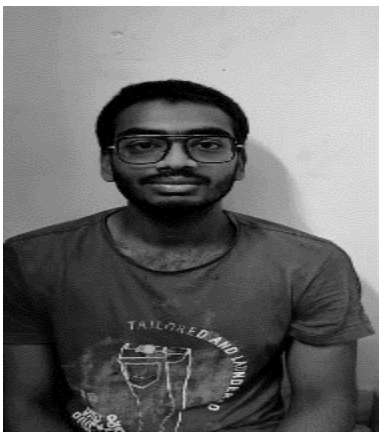


Figure 6. Face Dataset Extraction



Sample image that is used for training the module.

b. Data Preparation and Training for Face Recognition

The dataset is read and given the classification label for different person images. Later the scaling of the images is to set a fixed size. Then the images are converted to greyscale before feeding them to the layers of the CNN algorithm as shown in Figure 7.

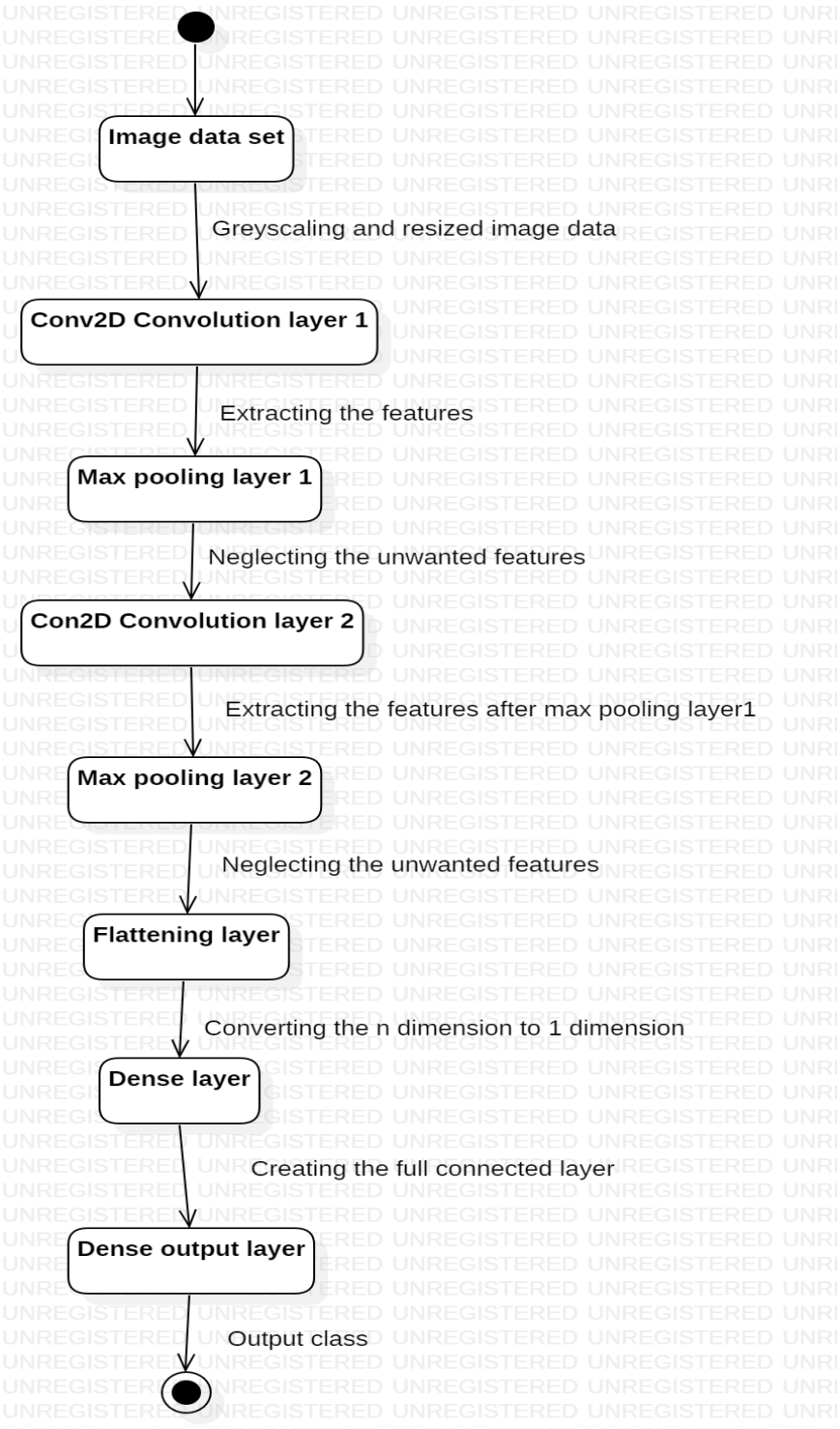


Figure 7. Face Recognition Model Layer

c. Data Prediction

When the user comes in contact with the camera, the face is captured, rescaled and grayscale image is given to the face recognition model.

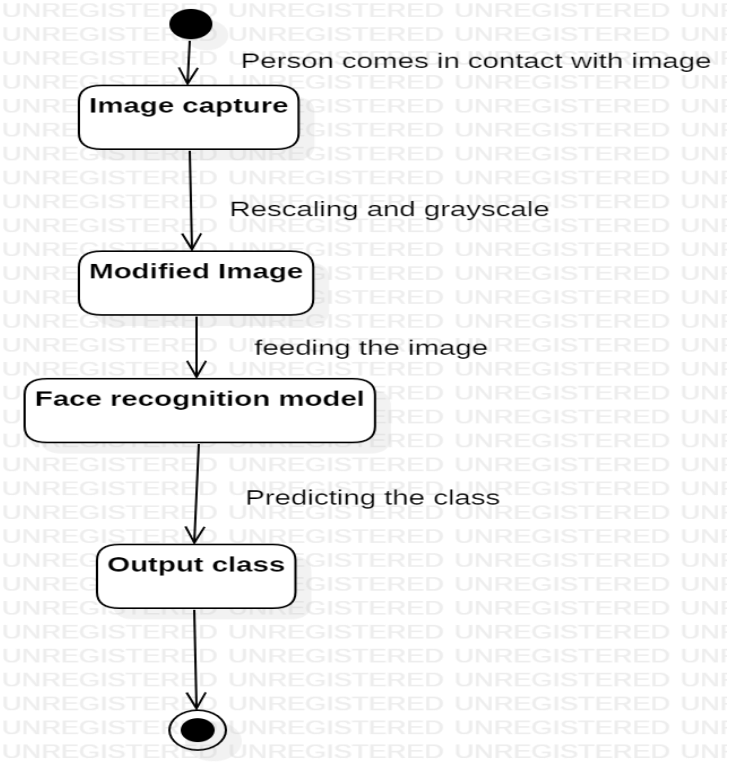


Figure 8. Face Prediction

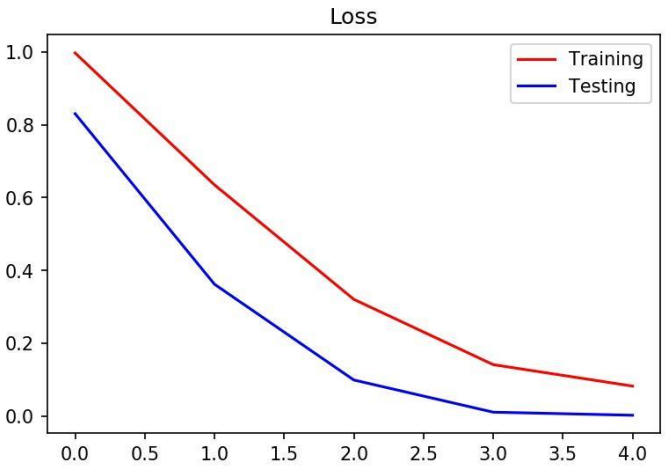


Figure 9. Graphical Representation of Face Recognition Model Training

iii. IBM Cloud Computation

The IBM cloud is a cloud service providing agent that provides many services for users to access them for free of cost, but of limited storage space. Usage of the cloud service in the project helps not only in enhancing the data security, but also in controlling the model from anywhere around the globe.The services likely come into role for this project development are

a. IBM Cloud

The cloud used to store the data text data from the business layer for cloud computation.

b. IBM Watson-Assistant

The service which finds the pattern from the data received by the ibm cloud and gives the simplified output text.

c. IBM Node-red

The service to decide the flow of the data and post the output data of Watson assistant to API.Later the data is used by the python code in the business layer.

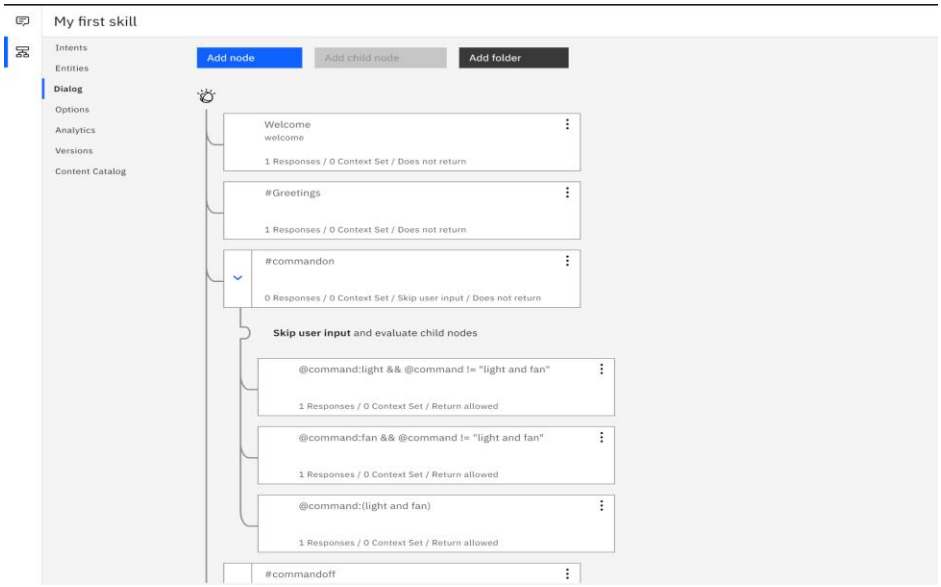


Figure 10. Watson Assistant Flow Dialog for Finding Pattern in Data

Conclusions

Things like choosing the best algorithm for the data processing and data prediction for features of speaker identification, speech recognition and smart surveillance are done in this piece of work. The Deep Learning Algorithm ANN taking the precision as evaluation metric helps in reducing the FP (False negative) i.e Predicting the unauthorized user as authorized user. The Deep Learning Algorithm CNN taking the loss as catogircal_crossentropy helps in predicting the different classes Usage of grayscale images for input to CNN algorithms helps in training the module in less time and attain efficiency of 96%. Usage of cloud service gives additional security to model and helps in transfer the data faster with REST calls.

The automation being an evolving technology that helps the users to accomplish their day-to-day targets can be enhanced by using the voice controlled automation. The module that performs the tasks based on the pattern of the speech given by the user helps voice automation technology to reach their true potential and makes it more

user friendly by not fixing the module only respond to limited no of commands. Making the voice automation module to take commands only from certain people helps the voice automation to expand their branches not only to household level but also to industrial level, where the tasks can be performed only after authentication.

References

1. Nath, P., & Pati, U. C. (2018). Low-cost android app based voice operated room automation system. *In 2018 3rd International Conference for Convergence in Technology (I2CT)*, 1-4.
2. Kinkiri, S., & Keates, S. (2020). Speaker Identification: Variations of a Human voice. *In 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 1-4.
3. Chayapathy, V., Anitha, G.S., & Sharath, B. (2017). IOT based home automation by using personal assistant. *In 2017 International Conference on Smart Technologies for Smart Nation (SmartTechCon)*, 385-389.
4. Rasanayagam, K., Kumarasiri, S.D.D.C., Tharuka, W.A.D.D., Samaranayake, N.T., Samarasinghe, P., & Siriwardana, S.E. (2018). CIS: an automated criminal identification system. *In 2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, 1-6.
5. Polyakov, E.V., Mazhanov, M.S., Rolich, A.Y., Voskov, L.S., Kachalova, M.V., & Polyakov, S.V. (2018). 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)*, 1-5.
6. Rajalakshmi, A., & Shahnasser, H. (2017). Internet of Things using Node-Red and alexa. *In 2017 17th International Symposium on Communications and Information Technologies (ISCIT)*, 1-4.
7. Sarker, G. (2012). An unsupervised learning network for face identification and subsequent localization. *In 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS)*, 632-635.
8. Wang, Y., & Chi, Z. (2016). System of wireless temperature and humidity monitoring based on Arduino Uno Platform. *In 2016 Sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC)*, 770-773.
9. Nan, E., Radosavac, U., Matić, M., Stefanović, I., Papp, I., & Antić, M. (2017). One solution for voice enabled smart home automation system. *In 2017 IEEE 7th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, 132-133.
10. Matić, M., Stefanović, I., Radosavac, U., & Vidaković, M. (2017). Challenges of integrating smart home automation with cloud based voice recognition systems. *In 2017 IEEE 7th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, 248-249.
11. Alshu' Eili, H., Gupta, G.S., & Mukhopadhyay, S. (2011). Voice recognition based wireless home automation system. *In 2011 4th international conference on mechatronics (ICOM)*, 1-6.

12. Pleshkova, S., Bekyarski, A., & Zahariev, Z. (2019). Reduced Database for voice commands recognition using cloud technologies, artificial intelligence and deep learning. *In 2019 16th Conference on Electrical Machines, Drives and Power Systems (ELMA)*, 1-4.

13. Saheaw, W., Jaiyen, S., & Hanskunatai, A. (2020). Thai Voice Recognition for Controlling Electrical appliances Using Long Short-Term Memory. *In 2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, 697-700.

14. Sravanthi, G., Madhuri, G., Sharma, N., Tiwari, A., Kashyap, A., & Suresh, B. (2018). Voice recognition application based home automation system with people counter. *In 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 574-578.

15. Mittal, Y., Toshniwal, P., Sharma, S., Singhal, D., Gupta, R., & Mittal, V.K. (2015). A voice-controlled multi-functional smart home automation system. *In 2015 Annual IEEE India Conference (INDICON)*, 1-6.

16. Lecouteux, B., Vacher, M., & Portet, F. (2011). Distant speech recognition for home automation: Preliminary experimental results in a smart home. *In 2011 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 1-10.