Thesis/Project No.: CSER-22-42.

Smart Home Automation Using Bangla Voice Command

By

Nohel Kumar Nath

Roll: 1607013

Pranto Sarker

Roll: 1607024

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna 9203, Bangladesh

27 March, 2022

Smart Home Automation Using Bangla Voice Command

By

Nohel Kumar Nath
Roll No. 1607013
&
Pranto Sarker
Roll No. 1607024

A thesis submitted in partial fulfillment of the requirements for the degree of
"Bachelor of Science in Computer Science and Engineering"

Supervisor:

Dr. M.M.A. Hashem

Professor
Department of Computer Science and Engineering
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh

Signature

Department of Computer Science and Engineering
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh

27 March, 2022

Acknowledgment

All the praise to the Almighty, whose blessing and mercy succeeded us to complete this thesis work fairly. After that, we humbly acknowledge the valuable suggestions, advice, guidance and sincere co-operation of Dr. M.M.A. Hashem, Professor, Department of Computer Science and Engineering, Khulna University of Engineering & Technology, under whose supervision this work was carried out. His intellectual advice, encouragement and guidance make us feel confident and scientific research needs much effort in learning and applying and need to have a broad view of problems from different perspectives. We would like to convey our heartiest gratitude to all the faculty members, officials and staff of the Department of Computer Science and Engineering as they have always extended their co-operation to complete this work. Last but not least, we wish to thank our friends and family members for their constant support.

-Authors

Abstract

Home automation is a famous and most used technology in the world. The Automation technology, life is getting simpler and easier in all aspects. In today's world Automatic systems are being preferred over manual system. With the rapid increase in the number of users of internet over the past decade has made Internet a part and parcel of life. Control basic home functions and features automatically through internet from anywhere around the world, an automated home is sometimes called a smart home. Numerous techniques are available today to build a home automation system There are many types of method such as inbuild voice command home automation system, wi-fi controlled voice command home automation, IoT based home automation, Bluetooth connected home automation etc. In this project the system introducing to machine learning where Bangla language data will be trained to improve the system. Most of the early research, many different types of method have been used to recognize the dataset at high accuracy. From many we will be working on MFCC (Mel Frequency Cepstral Coefficient) for feature extraction. Then our data will be trained on Convolutional Neural Network to gain the optimal accuracy in our model. There are also many features that can be extended further. To develop a Bangla short speech voice-command, accuracy is very important. The proposed system obtained a test accuracy of 96%. A lot of research works have been done to accomplish luxury and smart requirements while some threw light on the special needs for elderly and disabled etc.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

In our day-to-day life voice-controlled system has been never been much useful as today. Based on many voice-controlled environments like Amazon Alexa, Google Home are being used widely in many smart home-automation voice-controlled system. According to Google the rate of people whom uses voice-controlled device in their daily routine is about 72 percent[10]. A user-friendly voice-controlled environment is very easy to use and can make our life much comfortable. With voice command we can easily access to surrounding home appliance. New homes that are currently being built also are offering more smart home automation for homeowners. Smart home technology is revolutionizing the way we design our homes.

1.2 Motivation

Our goal is to build a Bangla voice-controlled home automation system which can be monitor/controlled the home appliance from anywhere.

The reason for this is that homeowners can increase the value of their property with the addition of technology. New homes that are currently being built also are offering more smart home automation for homeowners. Smart home technology is revolutionizing the way we design our homes.

1.3 Problem Statement

Voice control of your home automation system gives you voice access to any function of your home automation system. It can be very useful for people with physical limitations - such as the elderly and the physically challenged. Voice controlled systems are a huge deal with

quadriplegics and others with the same level of movement restrictions. Anyone confined to a wheelchair and without control of their arms can still easily control almost any system connected to a home automation controller.

1.4 Objectives

The objectives of this thesis are given below:

- To build a system in Bangla voice command system.
- To managing all of your home devices from one place.
- Remote control of home functions.
- To build a system for elderly disabled

1.5 Scope of the Work

There are a few automations system that done combinedly with custom Bangla voice command and implemented on home automation system. There are only a few contributions in Bangla speech recognition [4]. With custom trained voice data will help us to improve over other home automation system with high accuracy Bangla voice command.

1.6 Methodology

We proposed to develop a Bangla voice-controlled home automation system which can be monitor/controlled the home appliance. Our proposed method will be implemented on Raspberry Pi which will allow to take the Bangla voice command and connect to home-automation system [15].

The design of our smart home automation system is to easily access to the home appliances through Bangla voice command. The voice command will be given to the module and it will verify the Bangla voice command to execute the given instruction. The instruction will be carried out which is connected with the appliances.

1.7 Contribution

Voice recognition will be a key part of the future of communication. We have contributed a custom dataset in our Bangla voice command in our model. We have developed a strong language model and train the system to generate the acoustic models which could be used for Bangla voice recognition. Then the model will be implemented on a reliable automation system which is connected to multiple home appliance, which can be controlled through Bangla voice command.

1.8 Thesis Organization

Chapter 2: In this chapter, we reviewed some previous works related to Bangla voice-controlled home automation

Chapter 3: In this chapter, we described the System Analysis & Design of our proposed system and the tools we have used to develop it.

Chapter 4: In this chapter, we presented some theoretical prospects for developing our method.

Chapter 5: In this chapter, we demonstrated our proposed methodology and how the system works.

Chapter 6: In this chapter, we analyzed our system experimental result of our proposed methodologies.

Chapter 7: In this chapter, we summarized the whole thesis, discussed our future plan and concluded the thesis.

Chapter 2

Literature Review

In recent years, researchers have done a lot of contributions to developing an effective smart home automation system. There are some related works which have been done before for some related purpose. By going through some papers, we have come to acknowledge many kinds of techniques related to our work like system over view, hardware design and implementation etc. There are also many features that can be extended further. Numerous techniques are available today to build a home automation system. Kodali et al. [2] proposed many types of method such as inbuild voice command home automation system, wi-fi controlled voice command home automation, IoT based home automation, Bluetooth connected home automation etc.

2.1 Bangla Short Speech Commands Recognition

Ahmed et al. [8] proposed convolutional neural network (CNN) architectures have been designed to recognize bangla short speech commands. Mel-frequency cepstral coefficients (MFCC) features have been extracted from the audio files in one approach whereas only the raw audio files have been used in another CNN architecture. Lastly, a pre-trained model which is trained on a large English short speech commands data set has been fine-tuned by retraining on Bangla data set. Experimental results reveal that the MFCC model shows better accuracy in recognizing Bangla short speech commands where, surprisingly, the model predicting on raw audio data is very competitive. The models have shown proficiency in identifying single syllable words but encounter difficulties in recognizing multi-syllable commands. The convolutional neural networks (CNN) are a simple extension of the multi-layer perceptron model which can be considered as a diverse version of the standard neural networks.

2.2 Bengali Spoken Digit Classification

Sharmin et al. [9] proposed Bengali speech recognition system which is the process of Bengali spoken digit classification, but all of them had missed out on one or two influential parameters like dialects, gender or age-groups. Voice of people differs due to gender, dialects, and age.

This paper proposes a deep learning approach for classifying the Bengali spoken digits. It takes all parameters like dialects, gender, age-groups into account and the proposed approach acquires more than 98% accuracy using a convolutional neural network (CNN). People say the same word differently based on their gender, age, and dialect. So, collecting voice recordings of people with different Bengali dialects, different genders, different age-groups helped this dataset to become diversified. MFCC or Mel Frequency Cepstrum Coefficient is a method of feature extraction from audio. Convolutional Neural Network (CNN) has been used for feature learning and classification purposes. It was created using the "Keras" module in python. Train data was passed through the 2D-convolutional layer (Conv2D) with a filter size of 32, 48 and 120.

2.3 Zigbee based home automation system

Ahmed et al. [26] proposed the automation focuses on recognition of voice commands and uses low-power ZigBee wireless communication modules together with a microcontroller. Easy VR is used to teach the voice commands with its compatible voice recognition software package named "Easy VR Commander". Also, the voice command is tested to ensure accurate recognition. This is done by speaking it in the microphone and then the command spoken is indicated by the software on the screen. The Easy VR Voice Recognition Unit is a multi-purpose voice recognition module designed to make the voice recognition system robust, cost-effective and versatile. The ZigBee communication protocol is used for an efficient transmission of data from the ZigBee transmitter to the ZigBee receiver.

2.4 Wi-Fi based home automation system using cell phones

Majeed [1] proposed Wi-Fi based home automation system mainly consist three modules, the server, the hardware interface module, and the software package. The figure shows the system model layout. Wi-Fi technology is used by server, and hardware Interface module to communicate with each other. The same technology uses to login to the server web-based application. The server is connected to the internet, so remote users can access server web-based application through the internet using compatible web browser. Software of the latest

home automation system is split to server application software, and Microcontroller (Arduino) firmware. The Arduino software, built using C language, using IDE comes with the microcontroller itself. Arduino software is culpable for gathering events from connected sensors, then applies action to actuators and preprogramed in the server. Another job is to report the and record the history in the server DB. The server application software package for the proposed home automation system, is a web-based application built using asp.net. The server application software can be accessed from internal network or from internet if the server has real IP on the internet using any internet navigator supports asp.net technology. Server application software is culpable of, maintain the whole home automation system, setup, configuration. Server use database to keep log of home automation system components, we choose to use XML files to save system log.

2.5 Home Automation Using Raspberry Pi

Rafi and Farhan [23] proposed the sphere of Home Automation System has lately been evolving by researchers and students giving them a platform to run their system with low costs, Raspberry Pi, Arduino and other controllers. The end user of existing systems would want to add more of flexible amenities to run the application via android. Hence the proposed system is built with compatibility and security factors into consideration and is achieved working with android phones. The interface consists of servers and sensors easy to configure with a Wi-Fi card port which is to be inserted and acts as a web server. The system can be accessed by the web browser through a Smartphone as well as a local PC using the server IP. Wi-Fi hence used provides a secured connection between the devices. The system provides the end users to control their devices/appliances through Smartphone thereby making the system free from wired connections. The information will however be routed through the Wi-Fi and accordingly the raspberry pi will be configured in a relay circuit as per the request from the user. The notifications can also be read through mails created by algorithms on raspberry pi.

2.6 Scope of Improvement in Existing Works

All these works discussed above are some of the significant contributions related to Bangla speech command recognition and home automation system. But there still have scopes of improvement to these works. Most of the work have done separately and implemented separately. There are a few automations system that done combinedly with custom Bangla voice command and implemented on home automation system.

Chapter 3

System Analysis & Design

Systems development is systematic process which includes phases such as planning, analysis, design, deployment, and maintenance. Here, in this tutorial, we will primarily focus on −

- Systems analysis
- Systems design

3.1 Systems Analysis

It is a process of collecting and interpreting facts, identifying the problems, and decomposition of a system into its components.

System analysis is conducted for the purpose of studying a system or its parts in order to identify its objectives. It is a problem-solving technique that improves the system and ensures that all the components of the system work efficiently to accomplish their purpose.

Analysis specifies **what the system should do**.

3.2 Systems Design

It is a process of planning a new business system or replacing an existing system by defining its components or modules to satisfy the specific requirements. Before planning, you need to understand the old system thoroughly and determine how computers can best be used in order to operate efficiently.

System Design focuses on **how to accomplish the objective of the system**.

System Analysis and Design (SAD) mainly focuses on −

- Systems
- Processes
- Technology

3.3 Smart Home System Analysis & Design:

Since the dawn of history mankind has been trying to achieve comfort combined with simplicity, as the goal is so bright it still stumbles with lots of obstacles. A home is the place where people usually relax, spend quality time with their families and sometimes even work on their jobs. Building a home went through many phases, at the stone age it was just a cave but now at the 21st century a home might be just a small place or a huge apartment in a skyscraper that would cost millions. The idea of building a smart home is considered a luxury for now but actually in few decades most of the houses will probably be a smart home in a way or another depending on its complexity, efficiency, and cost.

3.4 Conceptual Smart Home System Design:

i. Problems and Solutions

As people lived in their homes a lot of problems popped up which needs to be resolved. Some of the problems that people face in ordinary homes are the waste of energy and other resources, security, and time saving. The smart home solves these problems in a very effective way. They can easily control their light, fan, door by some voice command. It is actually very time-saving. As for the resource and energy saving, lighting system is connected to sensors in order to automatically turn on the lights when there is movement and when in active for certain time the lights turn off in order to improve power consumption. Efficiency in the smart house is covered by easy control over all devices available in the home. The control of these devices is done from any portable smartphone or tablet by a special application.

ii. Initial Planning for Smart Home

As designers go through more reading in order to achieve the required level of smart house in complexity and technology, prototypes are reviewed as they extend the knowledge of the designer engineers in order to evolve new designs in order to meet the customer's requirements if it is a custom made smart home not a commercial of the shelf, as if it was COTS the design would be implemented as it is and the ability to change the design will not be available. A brainstorming sit-down between the designers and the customer is required in order to establish new ideas so that would evolve. Executive engineers are required to be present in this sit-down to approve that the discussed ideas are applicable and could be done in real- life. After the brain storming with the customer, executive engineers and the designers need to work together to ensure the integration between the components used and the software that will operate within the house. A prototype is produced to the customer for his approval.

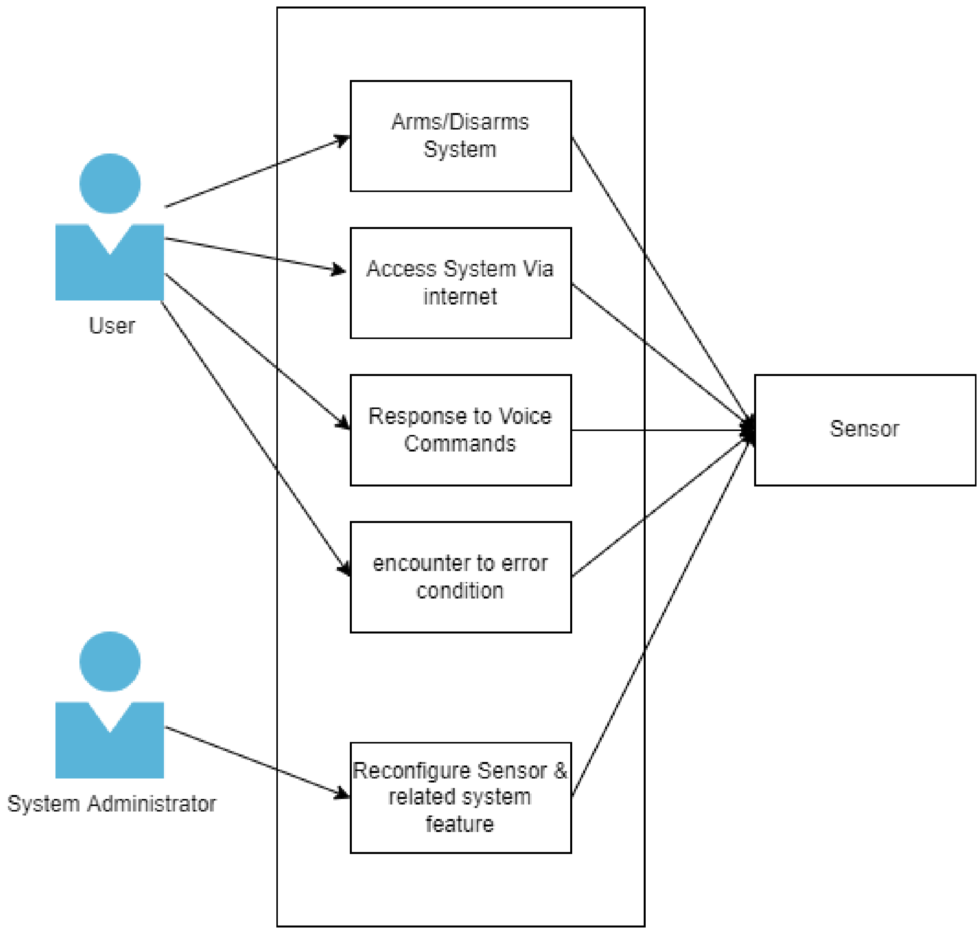

Fig 3.1: Smart Home Design

iii. Feasibility analysis for the Smart Home

In the world of smart houses different levels and technologies are used, as mentioned before there are different stages of complexity that can be used in designing a smart home, those different levels certainly come with different prices which will change the economical point of view of the customer, as he will decide their choice upon the allowed, needed, and the affordability of their home. As we all know the performance varies with the kind of level and complexity so the very famous and well known saying is applied "the more you pay, the more you get". The difference in the price could be due to different reasons, for example the quality of components used which will differ in the lifetime of system and its reliability. The price would also differ due to the different options available in the smart home. From figure 3.1, our design is chosen to be simple but effective, full of features but not expensive, the reason for this choice is to provide a good smart home with a low price. Although it's clear that our design is full of components but actually it is not highly priced, the economical view is taken in consideration to help in finding the most flexible and efficient home. A lot of smart homes are more complex than this design but we choose the most important and needed issues to solve. From the view of maintainability, our design is very easy to be maintained, there is no need to bring an expert each time a problem is found, any issue can be fixed in an easy and fast way, in other words people living in this smart home can fix any bug by themselves with no need to professional persons.

iv. Requirements for the operations of the Smart Home

In order to perform a well-designed and established smart house the quality of the equipment used must be superior quality that can handle a tough and around the clock usage of different members of the family. The location is a decisive variable in the requirements for operations as it determines most of the materials used and how to use it, for example the materials used in building a smart house in Alaska will certainly be different than materials used for building the same smart house in a different geographic location for example, the middle east. There are a lot of types for the equipment used are spread throughout the market, but we cannot say there is a perfect type because each person has his own need and has his own perfect view for his dream home but there is a certain criteria and certain limit that must be taken in consideration to achieve a good level for the equipment. The quantity of equipment depend on the home size and the person's need, in our design we choose limited quantity but with a good type because our idea is to achieve the goal with a reasonable but of course with great performance. Experts will take the lead in setting the basics of an efficient smart home. The number of experts needed is about 2-4 persons. Requirement for our smart home are as given below:

- The system must be able to register new members successfully.
- The system must be able to update control system.
- The system must allow users and administrators to view the system.
- The system must be updated with the latest hardware and software.
- The system must allow management to allocate administrators for personal monitoring.
- The system must allow users to provide feedback and management to review feedback.
- The system should ensure that the system are secure.
- The system should be optimized for performance.
- The system should not crash under heavy load.
- The system should avoid and reduce data redundancy.

Everything in life has an end and every product has a lifetime period, by estimating the lifetime of smart homes it can be said that the life time period of a smart home is around 5 years. A question here may be asked, why 5 years? , as its clear the smart home is full of electronic devices and even it is always updated and efficient it must have an end so considering this period it must be changed by the newest devices according to the latest technologies. According to our smart integrated wiring system the mission of changing devices and updating is so easy and simple so it will not take much time.

v. Smart Home maintenance and support Concepts

A good product does not only mean to get a high quality one but it also means to get a reliability one with a low probability of failure and when such a failure exist it is easy to maintain a support for it and find spare parts for the system which should be available and affordable to customers. This is important because everything including electronics and hardware have a specified life time after which it needs to be repaired or replaced. The better the maintenance policy the more it would attract clients to sign a contract from the designer's company. Another point of view when we talk about the customer support is how long does it cover, because time is the main aspect in this process as the products are affected by time and usage compared to any other variables. This task is simple and complex at the same time, the system is complex because it is full of equipment, wires, software and connections. The simplicity of the system occurs in its smart method of connecting and integrating all the components also in the multitasking of the devices that allows you to use your smart devices in operating any mission or task while sitting in your place. In order to accomplish this goal we need high personnel skills to perform well. Experts must be well trained on these devices and must be familiar with the latest updates and different technologies. As time goes on when smart homes spread all over the word there will be professionals specified in this field, and that's what we need to achieve our goal. Our equipment is divided into three parts: repairable, non-repairable and partially repairable. For example; the sensors used are non-repairable means that whenever the sensors is destroyed it cannot be repaired it is better to buy a new one . Finally the repairable devices which can be fully repaired like the smart phones and tablets which can repaired whenever a part of it has a problem (software or hardware). The environment of our system is not a big deal smart home can be constructed under any environmental condition and that is a positive point that makes the idea of this home fascinating. Whatever the temperature is our smart home will continue its mission to provide the comfortable dreaming home.

vi. Technical performance measures of Smart Home

There are different criteria which are necessary to be discussed among the technical performance measures (TPM). There is no doubt that the reliability, availability, affordability, and maintainability are these main aspects to be dealt with as the better they are the better the product is in the eyes of the customer. For the reliability our product's probability of failure must be as low as possible in order to tolerance the tough work. The availability of our system is high means that its available whenever you are, also the components choose to be used in the smart home must be available everywhere so we must choose our components from the market in order to compare the different type and measure there degree of availability. Another point of view in the field of the technical performance is the affordability and the difference between the availability and this point. Affordability is the ability to buy a product while availability is finding the product everywhere. In our system the affordability must be high in order to buy any missing product whenever is needed. Time processing is a very sensitive point because every person cares about time and needs his job to be done in the very little time. We tried as mentioned before to make this system the simplest way ever in order to save time also if we looked inside the system itself we will find that the components perform the tasks in a fast way. The final point to be discussed in this part is the maintainability or the ability of the products to be maintained, we are trying to reach the most maintained products ever in order to give a chance for the smart home to last for the longest time.

vii. Functional Analysis of the Smart Home

In this part the system is going to be analyzed, smart home system consists of a large number of subsystems and each subsystem contains a several number of components. The system is that huge general part than integrate all what's under it, our system is the smart home. The subsystems are as the following; intelligent home control, intelligent home entertainment, intelligent home comfort and smart home energy. Each subsystem of those has its own job for accomplishing the required goal of creating the most comfortable and creative home ever. Those subsystems contain components like: access control, integrated wiring, lightning control, electrical control, energy measurement and solar intelligent lightning. Those are the full components of our system that we hope to achieve our goal with, Depending on the experts and professional personnel also on the latest technological devices our system will lead all other systems in the last decades.

viii. The smart home trade-off analysis

As mentioned before there are different levels and different options ahead of the customers and designers and the choice is to be made under certain criteria such as: Commercial of the shelf and possible automation technology and custom made prototypes and many other support criteria and maintenance. These trade-offs are the original thoughts of the buyer when they go to choose their product.

ix. Smart Home specifications

For the specifications it is found that the most critical and valuable requirement of all is to convince the customer with meaningful and reasonable ways that a smart home is worth paying for. The maintenance policy should be satisfied in order to guarantee not only an affordable comforting home but a one that can be easily fixed or edited in case of a change in mind happens.

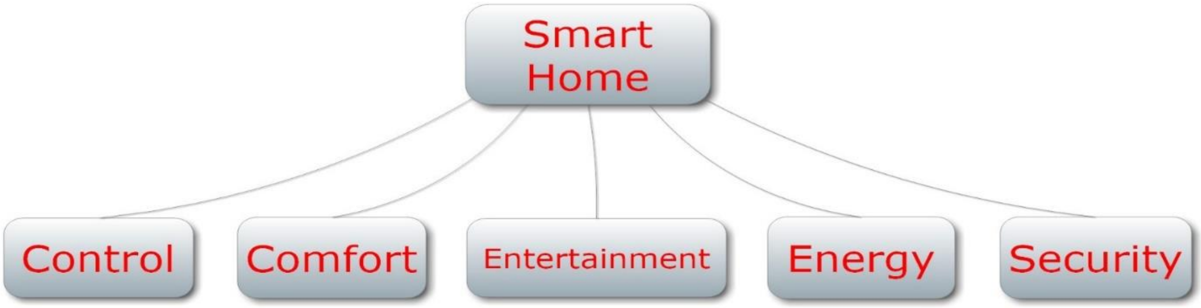

Fig 3.2: Smart Home specification

x. Conceptual design review

In conclusion, the problems of an ordinary home were identified and the solutions were figured out by the smart home. After the initial planning were done by brain storming ideas by the executive engineers and the clients, the feasibility were analyzed and dealt by the personnel who were experts in this field. The requirements for the operation of the smart home were picked up by the engineers to make sure that the facilities and the environment are good enough for such application. The support concept is decisive from the point of view of the client as the better it is the more it pushes him/her forward in trying such a technology as the smart home. The technical performance measures and the functional analysis of the smart home are combined and evaluated seriously since they are key points in the conceptual design of this system. The smart home tradeoff analysis is chosen and the specifications are over viewed a number of times until it is easy to see that the implementation of such a home is possible.

Chapter 4

Theoretical Consideration

In this chapter, we review some of the theoretical concepts which helped us to develop the methodology for our study.

4.1 Mel Frequency Cepstral Coefficient

4.1.1 Pre-emphasis

Pre-emphasis refers to filtering that emphasizes the higher frequencies. Its purpose is to balance the spectrum of voiced sounds that have a steep roll-off in the high-frequency region. For voiced sounds, the glottal source has an approximately −12 dB/octave slope. However, when the acoustic energy radiates from the lips, this causes a roughly +6 dB/octave boost to the spectrum. As a result, a speech signal when recorded with a microphone from a distance has approximately a −6 dB/octave slope downward compared to the true spectrum of the vocal tract. Therefore, pre-emphasis removes some of the glottal effects from the vocal tract parameters. The most commonly used pre-emphasis filter is given by the following transfer function

$$H(z) = 1 - bz^{-1} \tag{4.1}$$

where the value of b controls the slope of the filter and is usually between 0.4 and 1.0.

4.1.2 Frame blocking and windowing

The speech signal is a slowly time-varying or quasi-stationary signal. For stable acoustic characteristics, speech needs to be examined over a sufficiently short period of time. Therefore, speech analysis must always be carried out on short segments across which the speech signal is assumed to be stationary. Short-term spectral measurements are typically carried out over

20ms windows, and advanced every 10ms.Advancing the time window every 10ms enables the temporal characteristics of individual speech sounds to be tracked, and the 20ms analysis window is usually sufficient to provide good spectral resolution of these sounds, and at the same time short enough to resolve significant temporal characteristics. The purpose of the overlapping analysis is that each speech sound of the input sequence would be approximately centered at some frame. On each frame, a window is applied to taper the signal towards the frame boundaries. Generally, Hanning or Hamming windows are used. This is done to enhance the harmonics, smooth the edges, and to reduce the edge effect while taking the DFT on the signal.

4.1.3 DFT spectrum

Each windowed frame is converted into magnitude spectrum by applying DFT.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-j2\pi nk}{N}}; \qquad 0 \le k \le N-1 \tag{4.2}$$

where N is the number of points used to compute the DFT.

4.1.4 Mel spectrum

Mel spectrum is computed by passing the Fourier transformed signal through a set of band-pass filters known as Mel-filter bank. A Mel is a unit of measure based on the human ears perceived frequency. It does not correspond linearly to the physical frequency of the tone, as the human auditory system apparently does not perceive pitch linearly. The Mel scale is approximately a linear frequency spacing below 1 kHz and a logarithmic spacing above 1 kHz [4]. The approximation of Mel from physical frequency can be expressed as

$$f_{mel} = 2595 \log_{10}(1 + \frac{f}{700}) \tag{4.3}$$

where f denotes the physical frequency in Hz, and f_{mel} denotes the perceived frequency. Filter banks can be implemented in both time domain and frequency domain. For MFCC computation, filter banks are generally implemented in frequency domain. The center frequencies of the filters are normally evenly spaced on the frequency axis. However, in order to mimic the human ears perception, the warped axis, according to the nonlinear function given in, is implemented. The most commonly used filter shaper is triangular, and in some cases the Hanning filter can be found. The triangular filter banks with Mel frequency warping. The Mel spectrum of the magnitude spectrum $X(k)$ is computed by multiplying the magnitude spectrum by each of the of the triangular Mel weighting filters.

$$s(m) = \sum_{n=0}^{N-1} [|X(k)|^2 \; H_m(k)]; \qquad 0 \leq m \leq M-1 \tag{4.4}$$

where M is total number of triangular Mel weighting filters [5, 6]. $H_m(k)$ is the weight given to the k^{th}energy spectrum bin contributing to the m^{th} output band and is expressed as;

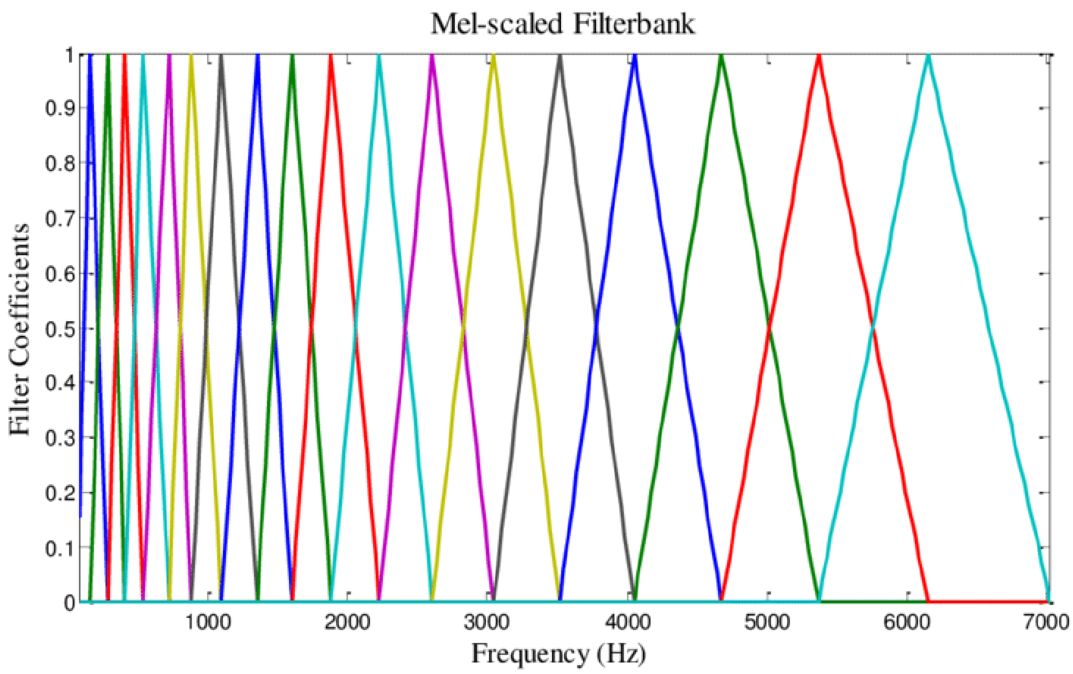

Fig 4.1: Mel-filter bank

$$H_m(k) = \begin{cases} 0\,, & k < f(m-1) \\ \dfrac{2(k-f(m-1))}{f(m)-f(m-1)}\,, & f(m-1) \le k \le f(m) \\ \dfrac{2(k-f(m-1))}{f(m)-f(m-1)}\,, & f(m) < k \le f(m+1) \\ 0\,, & k > f(m+1) \end{cases} \tag{4.5}$$

with m ranging from 0 to M − 1.

4.1.5 Discrete Cosine Transform (DCT)

Since the vocal tract is smooth, the energy levels in adjacent bands tend to be correlated. The DCT is applied to the transformed Mel frequency coefficients produces a set of cepstral coefficients. Prior to computing DCT, the Mel spectrum is usually represented on a log scale. This results in a signal in the cepstral domain with a quefrequency peak corresponding to the pitch of the signal and a number of formants representing low quefrequency peaks. Since most of the signal information is represented by the first few MFCC coefficients, the system can be made robust by extracting only those coefficients ignoring or truncating higher order DCT components. Finally, MFCC is calculated as.

$$c(n) = \sum_{n=0}^{M-1} \log_{10}(s(m)) \cos \frac{\pi n(m-0.5)}{M}\ ; \qquad n = 0,1,2, \dots, C-1 \tag{4.6}$$

where $c(n)$ are the cepstral coefficients, and C is the number of MFCCs. Traditional MFCC systems use only 8–13 cepstral coefficients. The zeroth coefficient is often excluded since it represents the average log-energy of the input signal, which only carries little speaker-specific information.

4.2 Convolutional Neural Network

The convolutional neural networks (CNN) are a simple extension of the multi-layer perceptron model which can be considered as a diverse version of the standard neural networks. In this section, we briefly discuss the architecture and other dimensions of traditional CNNs which are used mainly for speech recognition purposes

4.2.1 CNN Layer Architecture

A typical CNN for automatic speech recognition (ASR) introduces a different kind of network infrastructure compared to other artificial neural networks (ANN) and deep neural networks (DNN). However, traditional CNN consists of layers stacked together which are an input layer, a group of convolutional and pooling layers, several fully connected layers, and finally an output layer. The convolutional and pooling layers, followed by fully connected layers are the main differences of CNN compared to other neural networks, and this kind of special layer architecture has significant practical consequences in terms of speech recognition.

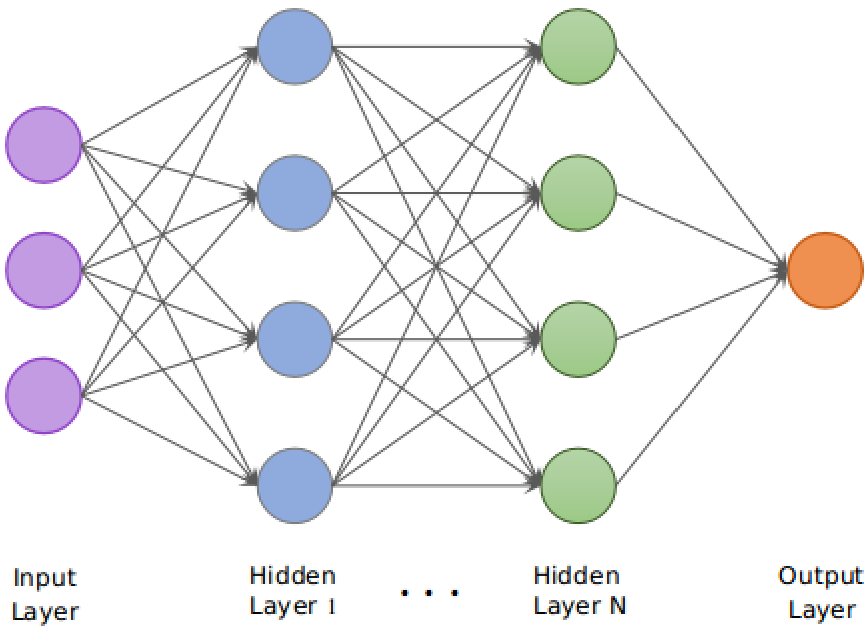

Fig 4.2: CNN Layer Architecture

4.2.2 Convolution Layer

A convolutional layer organizes hidden layer such a way that can take the advantages of input layer which is in form of a two-dimensional input data. Each hidden unit of convolutional layer processes only a small part of the whole input space rather than connecting to all the inputs coming from the previous input layer and it applies some arbitrary filters of an arbitrary dimension which result in feature map. From the feature map, CNN can understand local features of the data.

4.2.3 Pooling Layer

Another layer is connected with one convolutional layer which is called pooling layer. Pooling layer reduces the dimensionality of the extracted feature maps by applying a window of an arbitrary size which is called stride. It can extract either max, average or the sum of the windows. In this case, max pooling is used which extracted the highest values for each window in the feature map.

4.2.4 Batch normalization

Batch normalization is a layer that allows every layer of the network to do learning more independently. It is used to normalize the output of the previous layers. The activations scale the input layer in normalization. Using batch normalization learning becomes efficient also it can be used as regularization to avoid overfitting of the model. The layer is added to the sequential model to standardize the input or the outputs. It can be used at several points in between the layers of the model. It is often placed just after defining the sequential model and after the convolution and pooling layers.

4.2.5 Activation using rectified linear unit (ReLU)

ReLU (Rectified Linear Unit) Activation Function is the most frequently applied function in the CNN scenario. It transforms the input's entire values to positive integers. It is half rectified, when x<0 the functional value is zero and when x >=0 then function value is z. The values of the function and derivatives are monotonic. The key advantage of ReLU over the others is its lower computational load.

$$f(x)_{Relu} = \max\,(0, x) \tag{4.7}$$

here x is a real number in max (0, z), and it returns the maximum value.

Previously, logistic sigmoid and hyperbolic tangent have been used widely as non-linear activation functions in deep neural architectures like CNNs. But recently some alternative solutions have emerged and the application of Rectified Linear Units (ReLUs) is one of the most commonly used alternatives. Additionally, ReLU is the common alternative solution, since it has several advantages over typical activation functions which are faster computation and more efficient gradient propagation, biological plausibility and sparse activation structure.

4.2.6 Softmax Activation Function:

In multiclass classification, Softmax outperforms others. It is mostly used in output layer. Generally, Softmax activation is employed in the final layer to generate an output range of 0 to 1. The Softmax activation function computes the relative probabilities. It is quite similar to sigmoid activation function. It is mathematically represented as

$$\boldsymbol{softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{k} e x_i} \tag{4.8}$$

4.2.7 Optimizer Method:

The learning process has two primary issues: the first is the selection of a learning algorithm (optimizer); and the second is the usage of several enhancements (such as AdaDelta and momentum) along with the learning algorithm to improve the output. This model employs the Adam optimizer, which is an adaptive learning rate approach. It calculates different learning rates depending on the parameters. Adaptive moment estimation gets its name from the fact that it employs estimations of the first and second moments of the gradient to adapt the learning rate for each weight in the neural network. The n^{th} random variable moment is denoted as

$$M_n = E[X^n] \tag{4.9}$$

Where X stands for a random variable and M stands for the moment. The gradient of cost function in neural network can be regarded as a random variable. The mean is represented by the first moment, while the uncentered variance is represented by the second. Here uncentered variance means no subtraction of mean, while calculating the variance, Adam exponentially utilizes moving averages which is computed on gradient evaluated on a mini batch:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{4.10}$$

$$v_t = \beta_1 v_{t-1} + (1 - \beta_2) g_{t^2} \tag{4.11}$$

Where m and v represent moving averages, g stands for the gradient on current mini batch, and introduced the hyper parameters of the algorithm. Estimators expected values should be equal to the parameter being estimated. However, this isn't always the case, so the parameter becomes the expected value. If these statements are true, then it is an unbiased estimator. The general formula for moving averages is stated as equation dummy x.

$$m_t = (1 - \beta_1) \sum_{i=0}^{t} \beta_1^{t-1} g_i \tag{4.12}$$

The expected value of the moment, as well as the relationship between the first and expect ed moments, is discussed below:

$$E[m_t] = E[(1-\beta_1)\sum_{i=0}^{t}\beta_1^{t-1}g_i]$$

$$= E[g_i](1-\beta_1)\sum_{i=0}^{t}\beta_1^{t-1} + \vartheta$$

$$= E[g_i](1-\beta_1^t) + \vartheta \tag{4.13}$$

Here this equation shows the new method for calculating moving average. *G[i]* and *g[t]* are approximated so, it can be taken out from the sum and further it acts like an independent

variable. Because of approximation, error was emerged. In this scenario, Adam is a biased optimizer like many others. Correction of estimator is compulsory as it needs to obtain the expected value. The estimators' equations are noted below:

$$m_t^\wedge = \frac{m_t}{(1-\beta_1^t)} \tag{4.14}$$

$$v_t^\wedge = \frac{v_t}{(1-\beta_1^t)} \tag{4.15}$$

Now the individual learning rate will be calculated for each parameter using these moving averages with equation.

$$w_t = w_{t-1} - \frac{nm_t^\wedge}{\sqrt{m_t^\wedge}+\in} \tag{4.16}$$

here w is the model weights, n is step size. So, in this manner adam optimizer's weights are updated.

4.2.8 Loss function:

The loss function is the function that computes the distance between the current output of the algorithm and the expected output. It is used to optimize neural network parameters. It's a method to evaluate how your algorithm models the data. The loss functions are mean squared error and cross entropy. The mean squared error for regression tasks and cross entropy is used for classification tasks.

i. Mean squared error (MSE): It is the most commonly used loss function for regression. The loss is the mean overseen data of the squared differences between true and predicted values, or writing it as a formula.

ii. Cross entropy: Cross entropy loss function is an optimization function which is used in case of training a classification model which classifies the data by predicting the probability of whether the data belongs to one class or the other class. When fitting a neural network for classification, Keras provide the following three different types of cross entropy loss function:

• **Binary_crossentropy:** Used as a loss function for binary classification model. The binary_crossentropy function computes the cross-entropy loss between true labels and predicted labels.

• **Categorical_crossentropy:** Used as a loss function for multi-class classification model where there are two or more output labels. The output label is assigned one-hot category encoding value in form of 0s and 1. The output label, if present in integer form, is converted into categorical encoding using keras.utils to_categorical method.

• **Sparse_categorical_crossentropy:** Used as a loss function for multi-class classification model where the output label is assigned integer value (0, 1, 2, 3…). This loss function is mathematically same as the categorical_crossentropy. It just has a different interface.

4.2.9 Hyperparameters

Hyperparameters are various settings that are used to control the learning process. CNNs use more hyperparameters than a standard multilayer perceptron (MLP).

i. Kernel size

The kernel is the number of pixels processed together. It is typically expressed as the kernel's dimensions, e.g., 2x2, or 3x3.

ii. Padding

Padding is the addition of (typically) 0-valued pixels on the borders. This is done so that the border pixels are not undervalued (lost) from the output because they would ordinarily participate in only a single receptive field instance. The padding applied is typically one less than the corresponding kernel dimension. For example, a convolutional layer using 3x3 kernels would receive a 2-pixel pad on all sides.

iii. Stride

The stride is the number of pixels that the analysis window moves on each iteration. A stride of 2 means that each kernel is offset by 2 pixels from its predecessor.

iv. Filter size

Common filter sizes found in the literature vary greatly, and are usually chosen based on the data set. The challenge is to find the right level of granularity so as to create abstractions at the proper scale, given a particular data set, and without overfitting.

v. Dropout

Dropout is used as a regularization technique — it prevents overfitting by ensuring that no units are codependent. When we apply dropout to a neural network, we're creating a "thinned" network with unique combinations of the units in the hidden layers being dropped randomly at different points in time during training. Each time the gradient of our model is updated, we generate a new thinned neural network with different units dropped based on a probability hyperparameter p. Training a network using dropout can thus be viewed as training loads of different thinned neural networks and merging them into one network that picks up the key properties of each thinned network.

This process allows dropout to reduce the overfitting of models on training data. Applying a dropout layer enhances the overall result.

vi. Early stopping

Stop training automatically when a specific performance measure (eg. Validation loss, accuracy) stops improving.

Chapter 5

Methodology

5.1 Overview

We proposed to develop a Bangla voice-controlled home automation system which can be monitor/controlled the home appliance. The design of our smart home automation system is to easily access to the home appliances through Bangla voice command. The system will verify the Bangla voice command to execute the given instruction. The instruction will be carried out a module which is connected with the appliances. Considering our study, we want to develop the system introducing to machine learning where Bangla language data will be trained to improve the system. Our data will be trained on Convolutional Neural Network to gain the optimal accuracy in our model. Most of the early research many different types of method have been used to recognize the dataset at high accuracy.

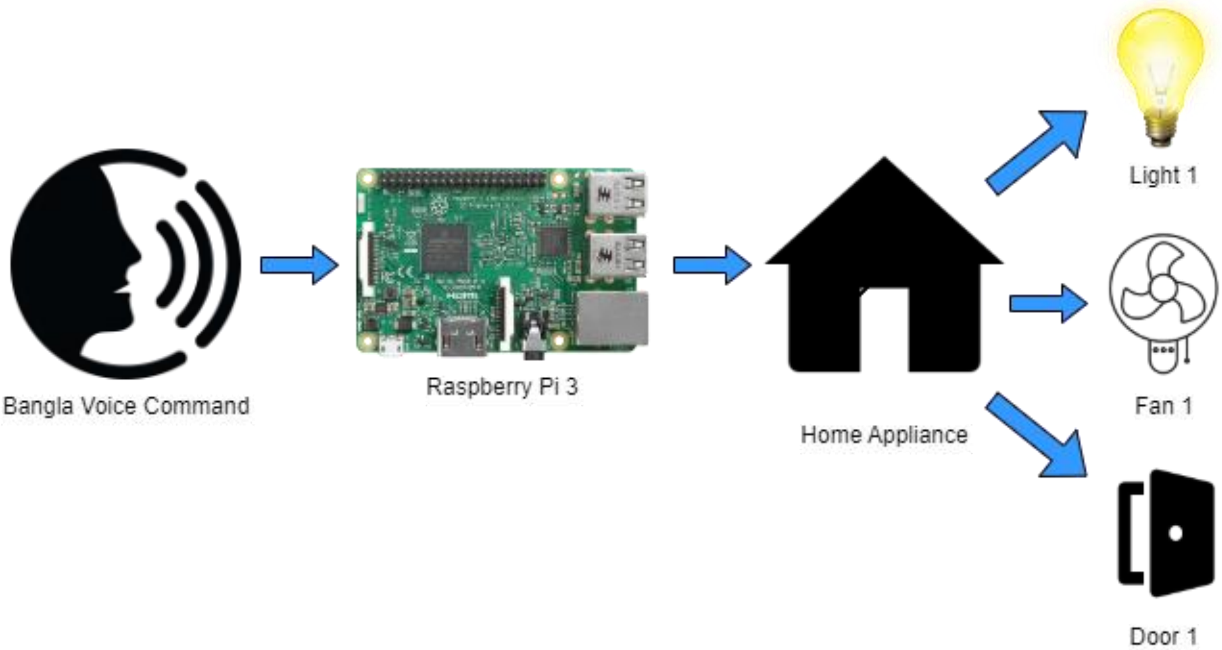

Fig 5.1: Overview of the system

5.2 Bangla Voice Command

Considering our study, we want to develop the system introducing to machine learning where Bangla language data will be trained to improve the system. Our data will be trained on Convolutional Neural Network to gain the optimal accuracy in our model. Most of the early research many different types of method have been used to recognize the dataset at high accuracy.

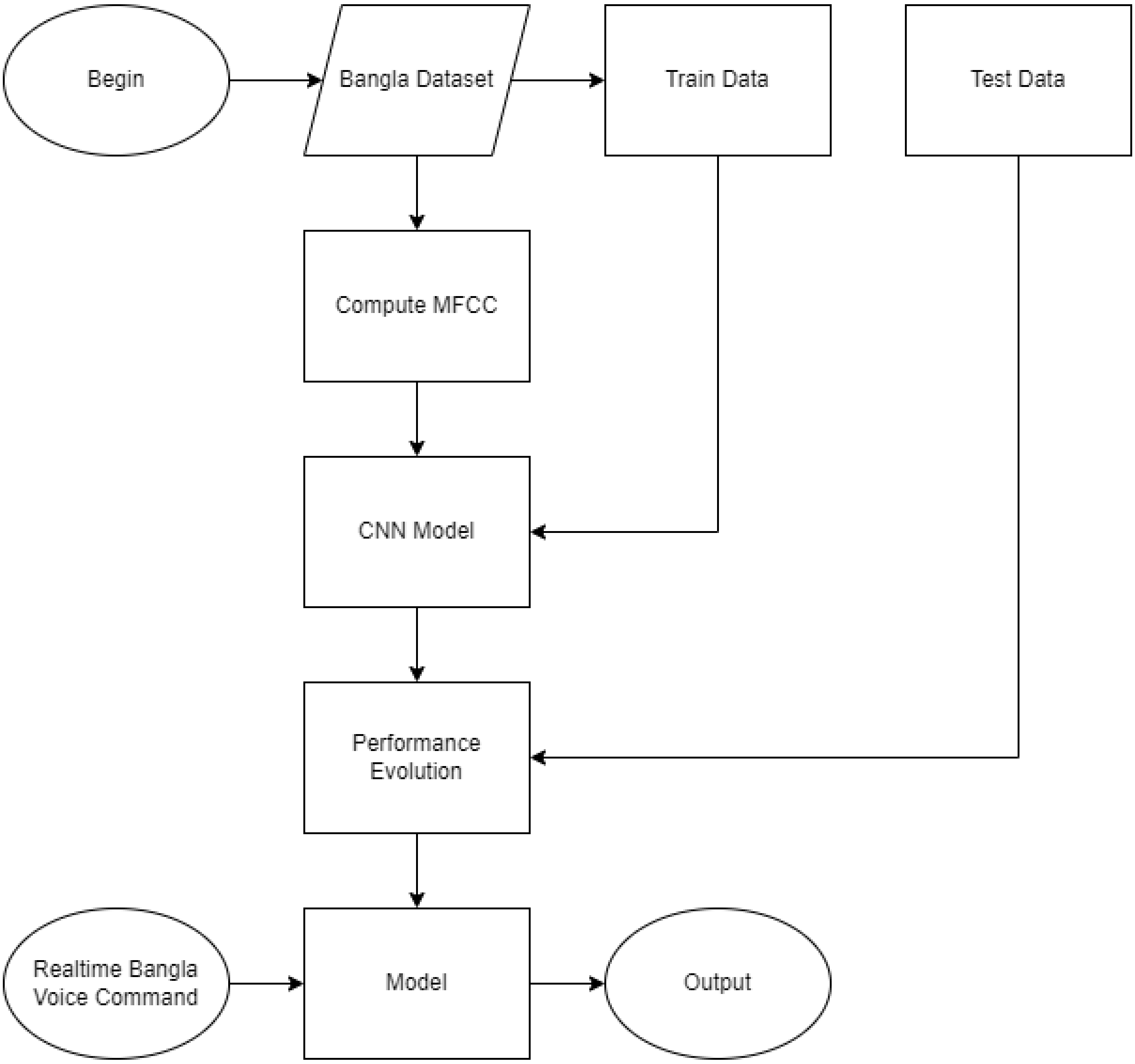

Fig 5.2: Proposed Bangla Voice Command CNN Model

5.3 Dataset Description

The resources for Bangla speech recognition are not widely available while there has been previous work on Bangla speech recognition, the dataset employed are not available publicly for research purpose, therefore we created our voice recognition dataset[9]. Our dataset containing 2835 audio file was created for the experiment. This audio file was converted to wav file Many friends were asked to give their voice recordings for the experiment. We recorded voice data with our microphone. The dataset contains total 10 command. All the text command of the wav files were in Bengali. The duration of each audio file is not more than 3 second. The sampling rate was 44080 Hz and bit rate was 320 kbps. Collecting voice recordings of our friends help this dataset to become diversified.

Table 4.1: Total number of Dataset files

Bengali Command	Total Audio Files
এক নাম্বার পাখা চালাও	311
এক নাম্বার পাখা বন্ধ করো	321
এক নাম্বার বাতি জ্বালানো হোক	299
এক নাম্বার বাতি নিভাও	299
দরজা খোলো	279
দরজা বন্ধ করো	285
দুই নাম্বার পাখা চালাও	266
দুই নাম্বার পাখা বন্ধ করো	298
দুই নাম্বার বাতি জ্বালাও	298
দুই নাম্বার বাতি নিভাও	180

5.4 Feature Extraction

From many we will be working on MFCC (Mel Frequency Cepstral Coefficient). MFCC will be using for sensitivity at appropriate frequencies by converting the conventional frequency to Mel Scale.

The Mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency.

Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip[27]. The difference between the cepstrum and the Mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum.

After getting the pre-processed data, we extracted the features by using Mel-frequency cepstral coefficients. The MFCC feature extraction splits the audio signal into short timestamps. After the splitting audio signal into the short timestamps, the periodogram estimate of the power spectrum has been calculated for each frame. Then, the Mel-filter bank is applied to the power spectrum and summed the filter energy. Furthermore, the logarithm of all filter bank energies has been calculated. Finally, the Discrete Cosine Transform (DCT) of the log filter bank energies are considered and keep the DCT coefficients. MFCC features were extracted from the audio files using the "librosa” module of python and appended into a python “numpy”-array. Then the data were carefully labelled as 0,1,2,3,4.. and etc.

5.5 Train Test Split

After feature extraction and labelling was completed, the dataset created was divided into “train” and “test” set. It was done using the test_train_split method from the “sklearn” module in python. The dataset was divided by maintaining 80:20 ratio; 80% data was used for the training dataset and 20% data was used for the test dataset.

5.6 Feature Learning and Classification using CNN

Convolutional Neural Network (CNN) has been used for feature learning and classification purposes. Fig 1. describes the precise CNN architecture that was used for this experiment. It was created using the “Keras” module in python[14].

Proposed model has 3 Convolution layers, 3 Batch Normalization layers, 3 max pooling layers, 1 flatten layer, 2 dense layers and 1 dropout layer. Description of this sequential model is as given below:

1) First 2D-convolution layer with a filter size of 16 and kernel size of (3,3). This layer uses Relu activation function.
2) One Batch Normalization layer.
3) One max pooling layer size of (3,3).
4) Second 2D-convolution layer with a filter size of 32 and kernel size of (3,3). This layer uses Relu activation function.
5) Second Batch Normalization layer.
6) Second max pooling layer size of (3,3).
7) Third 2D-convolution layer with a filter size of 64 and kernel size of (2,2). This layer uses Relu activation function.
8) Third Batch Normalization layer.
9) Second max pooling layer size of (2,2).
10) Then flatten layer is used.
11) One dense layer is used with softmax activation function and 64 nodes for the classification.
12) Dropout layer is introduced with 30% dropout rate.
13) Then, another dense layer is used with softmax activation function and 10 nodes for the classification.

Thus ends the Feature learning process.

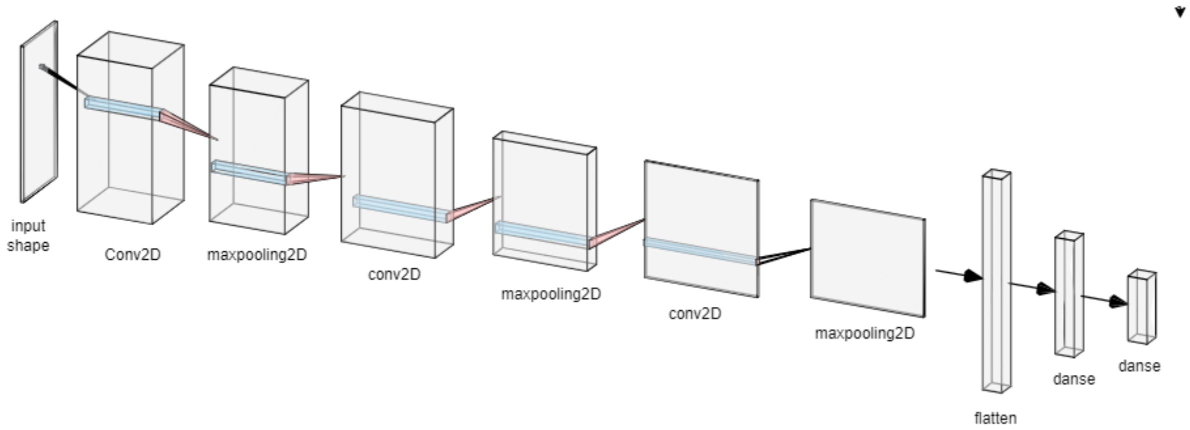

Fig 5.3: Architecture of the proposed approach using CNN

5.7 Hardware and Implementation

In this chapter we will discuss about hardware and implementation of our home automation system.

i. Component Details

The Raspberry Pi (Raspberry Pi Foundation, n.d.) is a credit-card-sized computer developed by Raspberry Pi Foundation, a UK-based charity. A Raspberry PI can be used in electronics projects and used like a computer to run software like spreadsheets, word processors, internet browsers and games.

Raspberry PI 3 Model B is used as the home client system. The system has a Quad Core 1.2GHz Broadcom BCM2837 64bit CPU, 1GB RAM, BCM43438 wireless LAN and Bluetooth Low Energy. A Node.js based web server acts as the home client and is connected directly to the internet. Firebase enables the raspberry to send status updates and receive commands in Realtime. The GPIO pins on the raspberry pi can be used directly to control electrical equipment via relays. Or it can be used to control additional microcontroller, setup in switch boards to connect

to other electrical components. The Raspberry PI used in this project has the following specifications:

- Quad Core 1.2GHz Broadcom BCM2837 64bit CPU

- 1GB RAM
- BCM43438 wireless LAN and Bluetooth Low Energy (BLE) on board
- 14
- 40-pin extended GPIO
- 4 USB 2 ports
- 4 Pole stereo output and composite video port
- Full size HDMI
- CSI camera port for connecting a Raspberry Pi camera
- DSI display port for connecting a Raspberry Pi touchscreen display
- Micro SD port for loading your operating system and storing data
- Upgraded switched Micro USB power source up to 2.5A

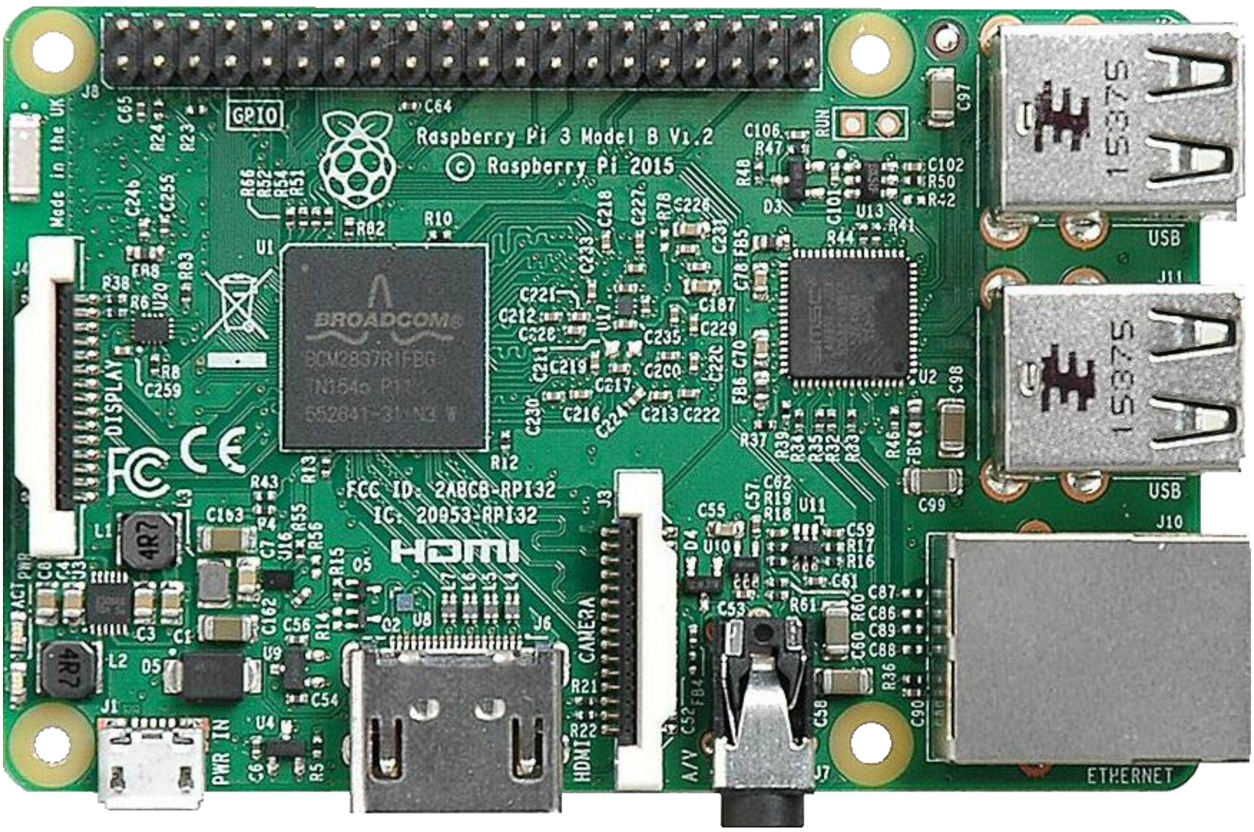

Fig 5.4: Raspberry Pi 3 Model B

ii. Model Implementation

In order to implement our model to raspberry pi we have to go through some steps. The implementation steps are given below:

I. First, we have to install a suitable OS for raspberry pi.
II. Then we have to connect the Raspberry Pi with the internet. With IP address we have connected raspberry pi with Wi-Fi.
III. Then we load our python model to our raspberry pi.
IV. Virtual environment must set to for python to run our model.
V. Raspberry Pi must be connected to Wi-Fi to desired home appliance where we have initially connected light 1, light 2, fan 1, fan 2, door 1 and door 2.

iii. Communication Design

For communication we have used a microphone which is connected to Raspberry Pi. As there is no microphone port in Raspberry Pi, we have used an USB sound card to connect our microphone.

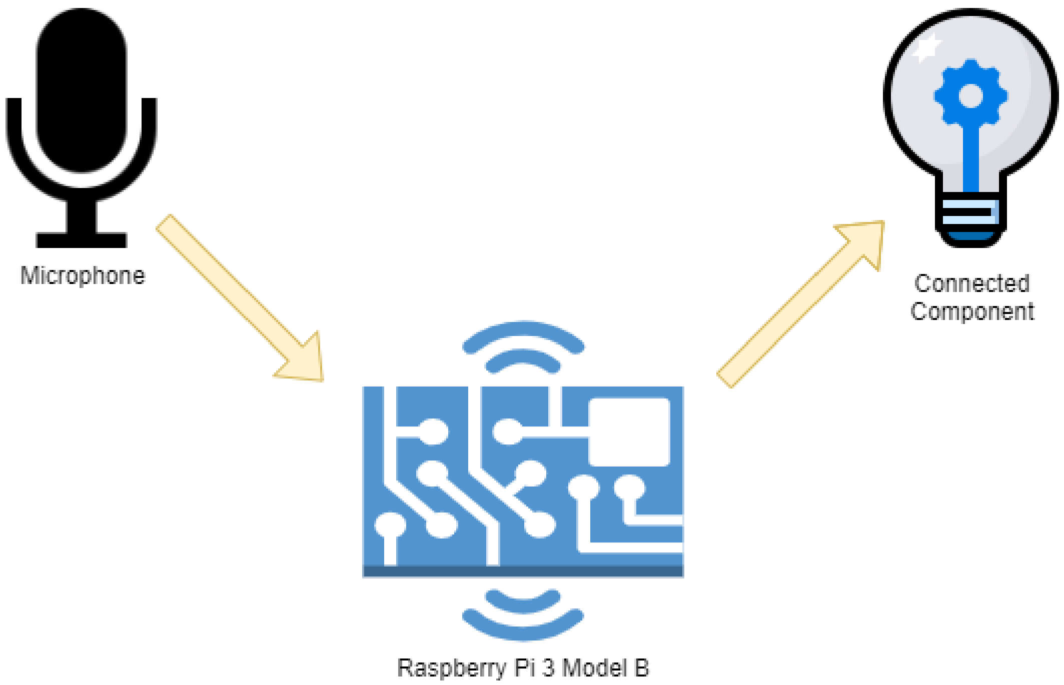

Fig 5.5: Communication Design with Raspberry Pi

iv. Circuit Design

The circuit has been implemented for demonstration.

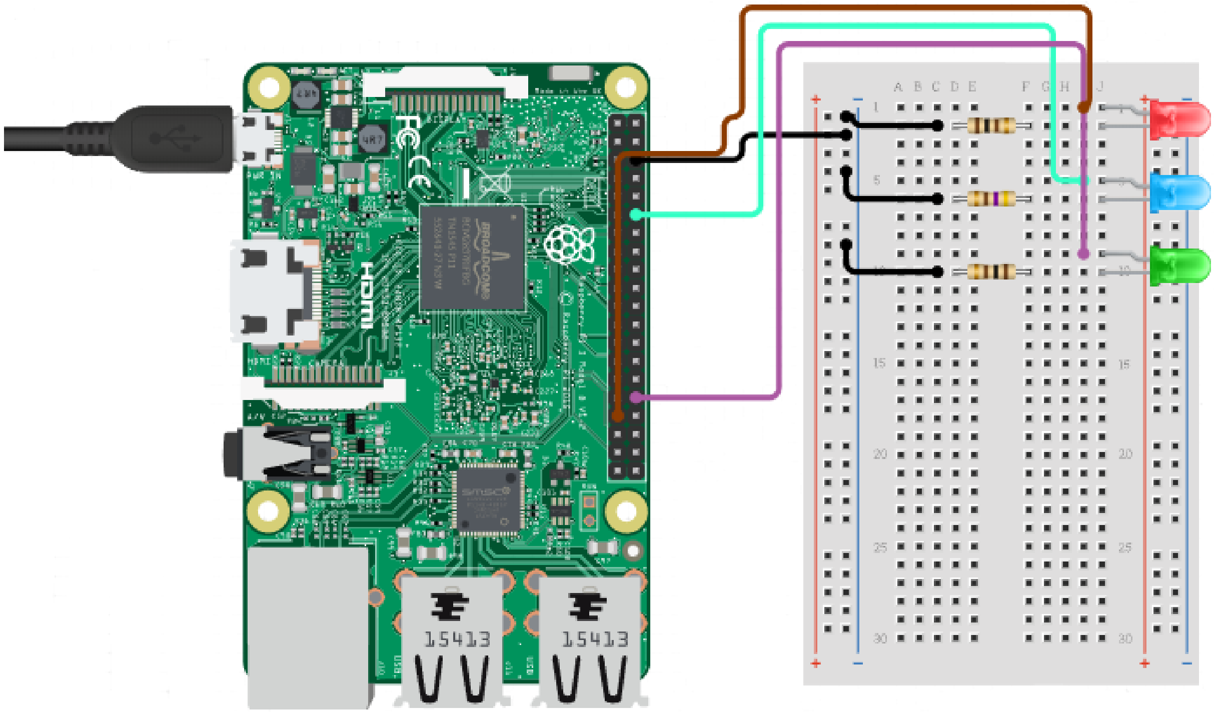

Fig 5.6: Circuit Design Implementation

The circuit shows the connection for a light and a switch. The light can be controlled using Bangla Voice command and the switch can be used to toggle the light and update the Android Application

Chapter 6

Experimental Result

At the stage of implementation of this thesis, we have analyzed our experiment results of our dataset on CNN. We have our visualization in many ways and have the analogy of the models below.

6.1 Experimental Setup

In this chapter, useful information about the hardware used in building the dataset and training and testing the proposed model are discussed. The experiments and analysis processes are done on a computer with intel core i5 processor having 4 cores with each core having 1.19 GHz Speed. Also, the system had 8 GB of Ram. The Graphics card that we have used is intel UHD graphics. The integrated development environment (IDE) that we have used is Anaconda, pycharm, spider, google co-lab and Kaggle.

6.2 Dataset

The goal of our study is to recognize the Bangla command from the audio that have been recorded in the microphone. Due to lack of open-source datasets, we have created a dataset. The dataset contains 2835 audio files. The dataset consists of 10 Bangla voice. Each command consists of almost 280 audio files. The audio files were recorded using our mobile microphone. The audio files were taken from the students of the department of computer science and engineering (CSE) of the Khulna University of Engineering & Technology (KUET), Khulna, Bangladesh.

6.3 Result analysis

Table 6.1 shows the accuracy of our dataset.

Table 6.1: Result analysis

Model	Number Of Epoch	Accuracy	Loss	Number Of Parameters
CNN	48	96.5%	13.3%	145,290

6.4 Confusion Matrix

The confusion matrix is also known as an error matrix. It is a table that measures the performance of a classification model where the true value of given inputs is known. For a classification problem, it summarizes the prediction result. The confusion matrix shows the prediction with the prearranged label. This matrix consists of True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). Table 6.2 shows a confusion matrix.

Table 6.2: Confusion Matrix table

	Actually Positive (1)	Actually Negative (0)
Predictive Positive (1)	True Positive	False Positive
Actually Negative (0)	False Negative	True Negative

True Positive (TP): True positives are the cases when the actual class of the data point was 1(True) and the predicted is also 1(True).

True Negative (TN): True negatives are the cases when the actual class of the data point was 0(False) and the predicted is also 0(False).

False Positive (FP): False positives are the cases when the actual class of the data point was O(False) and the predicted is 1(True).

False Negative (FN): False negatives are the cases when the actual class of the data point was 1(True) and the predicted is 0(False).

Accuracy: Accuracy is a measure that tells how much accurate the result is. It is expressed in:

$$Accuracy = \frac{True\ Positive + True\ Nagetive}{True\ Positive + True\ Nagetive + Flase\ Positive + Flase\ Nagetive} \quad (6.1)$$

Precision: Precision gives a measure that tells how much data objects are correctly and positively classified out of the all positively predicted data objects. It is expressed as:

$$Precision = \frac{True\ Positive}{True\ Positive + Flase\ Positive} \quad (6.2)$$

Recall: Recall gives a measure that tells how much data objects are correctly and positively classified out of the all actually positive data objects. It is expressed as:

$$Recall = \frac{True\ Positive}{True\ Positive + Flase\ Nagetive} \quad (6.3)$$

F-measure: The F measure (FI score or F score) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6.4)$$

Table 6.3 describes the results found using the CNN model for different sizes of datasets

Data Size	Accuracy	Precision	Recall	F1-score
250	38.30%	38.00%	38.00%	38.00%
953	90.10%	90.00%	90.00%	90.00%
2853	96.50%	96.00%	96.00%	96.00%

The following section provide the differences for different size of dataset:

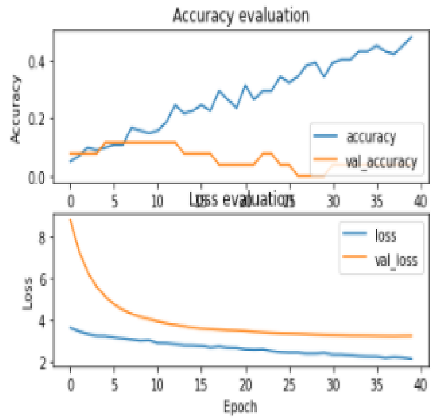

Fig 6.1: 250 Dataset Accuracy Evaluation and Loss Evaluation

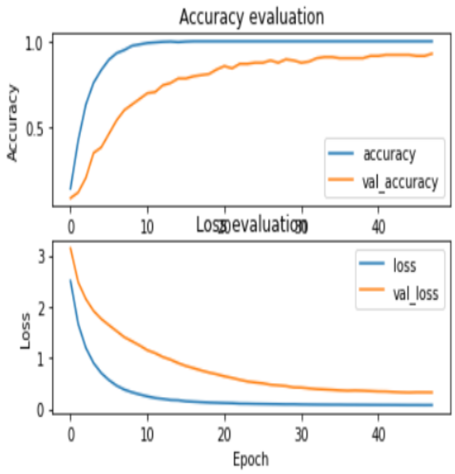

Fig 6.2: 953 Dataset Accuracy Evaluation and Loss Evaluation

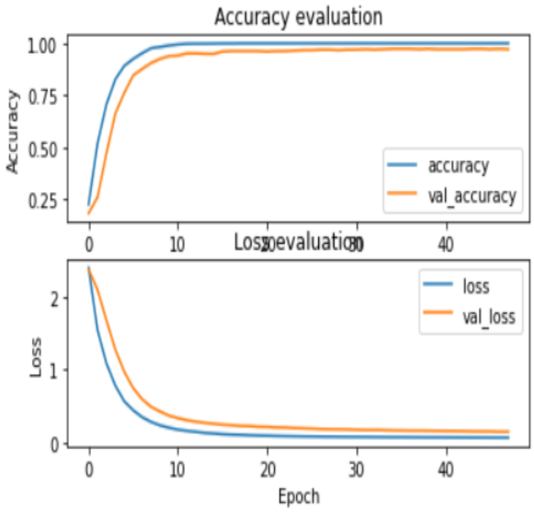

Fig 6.3: 2853 Dataset Accuracy Evaluation and Loss Evaluation

Table 6.4 shows the results of each command's precision, recall and f1-score found by using the CNN model based on their label are given below:

Command	Precision	Recall	F1-score
এক নাম্বার বাতি নিভাও	0.957	1.000	0.978
এক নাম্বার বাতি জ্বালানো হোক	0.985	0.971	0.978
দুই নাম্বার বাতি নিভাও	1.000	0.968	0.984
এক নাম্বার পাখা বন্ধ করো	0.981	1.000	0.990
এক নাম্বার পাখা চালাও	1.000	0.982	0.991
দরজা খোলো	1.000	1.000	1.000
দুই নাম্বার পাখা চালাও	1.000	0.962	0.981
দরজা বন্ধ করো	0.982	0.982	0.982
দুই নাম্বার বাতি জ্বালাও	0.966	0.983	0.974
দুই নাম্বার পাখা বন্ধ করো	1.000	1.000	1.000

Table 6.5: Result of test Bengali voice command [26]

Command	Total Number of Command	Accurate	Error
এক নাম্বার বাতি নিভাও	10	8	2
এক নাম্বার বাতি জ্বালানো হোক	10	6	4
দুই নাম্বার বাতি নিভাও	10	9	1
এক নাম্বার পাখা বন্ধ করো	10	8	2
এক নাম্বার পাখা চালাও	10	9	1
দরজা খোলো	10	9	1
দুই নাম্বার পাখা চালাও	10	8	2
দরজা বন্ধ করো	10	9	1
দুই নাম্বার বাতি জ্বালাও	10	7	3
দুই নাম্বার পাখা বন্ধ করো	10	7	3

The results show that more than 80% of success rate has been reached in a normal environment while Bangla voice commands were given to the system. The system works in smooth order in a normal environment. The result provides a clear idea to make command with a clear accent as the pre-stored command.

Chapter 7

Conclusion

7.1 Summary

Home automation system achieved great popularity in the last decades and it increases the comfort and quality of life. It may have seemed like an invention of the far distant future, but the reality is we have gotten closer than ever. While industry has much more progress than ever, there is still lots of room for improvement.

In this project the system introducing to machine learning where Bangla language data will be trained to improve the system. Most of the early research, many different types of method have been used to recognize the dataset at high accuracy. From many we will be working on MFCC (Mel Frequency Cepstral Coefficient) for feature extraction. Then our data will be trained on Convolutional Neural Network to gain the optimal accuracy in our model. The proposed system obtained a test accuracy of 96%. Then we will load our CNN model into our Raspberry Pi. Using Raspberry pi module, we can manage to connect over wi-fi in one module that will be connected to all home appliance like light, fan, door etc. User will give command over voice to the module that turn or control the appliance.

7.2 Limitations

Every work has some limitations. The limitations are:

1. All audio files are recorded in 3 sec. So, user command will have to be in exact 3 sec. otherwise it won't work properly.
2. In Realtime environment there are sometime limitation where our microphone straggle to take input voice.

3. The current technological limitations limit the ability to control all and any types of appliances. Since for our home and office appliances do not have built in micro controllers, controlling this electronic equipment cannot be possible.
4. We can only turn on or off most of the appliances.
5. There are no security system in our home automation system.
6. Our model has implemented on DC connection only.

Non-technical limitations include slow comprehension of Smart and Automated Equipment. Besides these, people are usually reluctant to steep changes.

7.3 Future Work

The current project can be extended in various ways. One such solution would be to integrate Alexa Voice Service (Amazon.com, n.d.) by Amazon.com, Inc. After integrating with Alexa Voice Service, the system can be operated using Amazon Echo Devices and other devices that have implemented the service. Control of electronic components operated by voice would be a marvelous solution for physically disabled people. Another possible extension could be the implementation of automation in factories and manufacturing plants.

Bibliography

[1] Rizwan Majeed, Nurul Azma Abdullah, Imran Ashraf, Yousaf Bin Zikria, Muhammad Faheem Mushtaq, Muhammad Umer, "An Intelligent, Secure, and Smart Home Automation System", *Scientific Programming*, vol. 2020, Article ID 4579291, 14 pages, 2020.

[2] R. K. Kodali, V. Jain, S. Bose and L. Boppana, "IoT based smart security and home automation system," *2016 International Conference on Computing, Communication and Automation (ICCCA)*, 2016, pp. 1286-1289, doi: 10.1109/CCAA.2016.7813916.

[3] H. Zunair, N. Mohammed and S. Momen, "Unconventional Wisdom: A New Transfer Learning Approach Applied to Bengali Numeral Classification," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8554435.

[4] Sumit Kumar Ghanty, Soharab Hossain Shaikh and N. Chaki, "On recognition of spoken Bengali numerals," *2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, 2010, pp. 54-59, doi: 10.1109/CISIM.2010.5643692.

[5] J. Islam, M. Mubassira, M. R. Islam and A. K. Das, "A Speech Recognition System for Bengali Language using Recurrent Neural Network," *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, 2019, pp. 73-76, doi: 10.1109/CCOMS.2019.8821629.

[6] C. Saha, R. H. Faisal and M. M. Rahman, "Bangla Handwritten Digit Recognition Using an Improved Deep Convolutional Neural Network Architecture," *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, pp. 1-6, doi: 10.1109/ECACE.2019.8679309.

[7] Ahammad, Khalil & Rahman, Md. Mahfuzur. (2016). Connected Bangla Speech Recognition using Artificial Neural Network. International Journal of Computer Applications. 149. 38-41. 10.5120/ijca2016911568.

[8] S. Ahmed Sumon, J. Chowdhury, S. Debnath, N. Mohammed and S. Momen, "Bangla Short Speech Commands Recognition Using Convolutional Neural Networks," *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8554395.

[9] Riffat Sharmin, Shantanu Kumar Rahut, Mohammad Rezwanul Huq, "Bengali Spoken Digit Classification: A Deep Learning Approach Using Convolutional Neural Network," Procedia Computer Science, Volume 171, 2020, Pages 1381-1388, ISSN 1877-0509.

[10] S. Dey, A. Roy and S. Das, "Home automation using Internet of Thing," *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2016, pp. 1-6, doi: 10.1109/UEMCON.2016.7777826.

[11] D. H. Stefanov, Z. Bien and W. C. Bang 2004, "The smart house for older persons and persons with physical disabilities: Structure, technology arrangements, and perspectives," In IEEE Trans. Neural Syst. Rehabil. Eng., Vol. 12, No. 2, pp. 228– 250

[12] H. Bharathi, U. Srivani, M. D. Azharudhin, M. Srikanth and M. Sukumarline, "Home automation by using raspberry Pi and android application," *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, 2017, pp. 687-689, doi: 10.1109/ICECA.2017.8212754.

[13] M. S. Mahamud, M. S. R. Zishan, S. I. Ahmad, A. R. Rahman, M. Hasan and M. L. Rahman, "Domicile - An IoT Based Smart Home Automation System," *2019 International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)*, 2019, pp. 493-497, doi: 10.1109/ICREST.2019.8644349.

[14] M. I. R. Shuvo, S. Akib Shahriyar and M. A. H. Akhand, "Bangla Numeral Recognition from Speech Signal Using Convolutional Neural Network," *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2019, pp. 1-4, doi: 10.1109/ICBSLP47725.2019.201540.

[15] J. Rahman Saurav, S. Amin, S. Kibria and M. Shahidur Rahman, "Bangla Speech Recognition for Voice Search," *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2018, pp. 1-4, doi: 10.1109/ICBSLP.2018.8554944.

[16] M. A. A. Amin, M. T. Islam, S. Kibria and M. S. Rahman, "Continuous Bengali Speech Recognition Based On Deep Neural Network," *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, pp. 1-6, doi: 10.1109/ECACE.2019.8679341.

[17] R. S. Raju, P. Bhattacharjee, A. Ahmad and M. S. Rahman, "A Bangla Text-to-Speech System using Deep Neural Networks," *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2019, pp. 1-5, doi: 10.1109/ICBSLP47725.2019.202055.

[18] U. Singh and M. A. Ansari, "Smart Home Automation System Using Internet of Things," *2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC)*, 2019, pp. 144-149, doi: 10.1109/PEEIC47157.2019.8976842.

[19] M. Sági, D. Mijic, D. Milinkov and B. Bogovac, "Smart home automation," *2012 20th Telecommunications Forum (TELFOR)*, 2012, pp. 1512-1515, doi: 10.1109/TELFOR.2012.6419507.

[20] M. Al-Kuwari, A. Ramadan, Y. Ismael, L. Al-Sughair, A. Gastli and M. Benammar, "Smart-home automation using IoT-based sensing and monitoring platform," *2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018)*, 2018, pp. 1-6, doi: 10.1109/CPE.2018.8372548.

[21] B. Mustafa, M. W. Iqbal, M. Saeed, A. R. Shafqat, H. Sajjad and M. R. Naqvi, "IOT Based Low-Cost Smart Home Automation System," *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2021, pp. 1-6, doi: 10.1109/HORA52670.2021.9461276.

[22] T. Chaurasia and P. K. Jain, "Enhanced Smart Home Automation System based on Internet of Things," *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2019, pp. 709-713, doi: 10.1109/I-SMAC47947.2019.9032685.

[23] Ar Rafi, Farhan. (2018). Application of Android and Raspberry Pi for Home Automation.

[24] Faroom, Saeed & Ali, Muhammad & Yousaf, Sheraz & Deen, Shamsa. (2018). Literature review on home automation system for physically disabled peoples. 1-5. 10.1109/ICOMET.2018.8346397.

[25] Elshafee, Ahmed & Alaa Hamed, Karim. (2012). Design and Implementation of a WiFi Based Home Automation System. WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY. 2177.

[26] Mohammed Arif Hasan Chowdhury, Jeenat Sultana. Bangla Voice Recognition Based Home Automation for Elderly and Disabled Patient (https://www.academia.edu/37475623/IRJET_Bangla_Voice_Recognition_Based_Home_Automation_for_Elderly_and_Disabled_Patient)

[27] Purkaystha, Bishwajit & Nahid, Md Mahadi & Islam, Md Saiful. (2019). End-to-End Bengali Speech Recognition using DeepSpeech.