A PROJECT REPORT ON

INTERACTIVE SYSTEM FOR PRODUCT PURCHASE THROUGH GESTURE AND VOICE

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

Gaurav Zanpure	**Exam No: B150314352**
Bhavik Ransubhe	**Exam No: B150314320**
Adesh Oak	**Exam No: B150314299**
Rithvik Poojary	**Exam No: B150314323**

Under The Guidance of

Prof. Dattatray Modani

DEPARTMENT OF COMPUTER ENGINEERING

P.E.S MODERN COLLEGE OF ENGINEERING,

PUNE-411005

SAVITRIBAI PHULE PUNE UNIVERSITY

2021 -2022

Progressive Education Society's
Modern College of Engineering
Department of Computer Engineering
Shivajinagar, Pune - 411005

CERTIFICATE

This is to certify that the project report entitled

INTERACTIVE SYSTEM FOR PRODUCT PURCHASE THROUGH GESTURE AND VOICE

Submitted By

Gaurav Zanpure	**Exam No: B150314352**
Bhavik Ransubhe	**Exam No: B150314320**
Adesh Oak	**Exam No: B150314299**
Rithvik Poojary	**Exam No: B150314323**

is a bonafide work carried out by Students under the supervision of **Prof. Dattatray Modani and it is submitted towards partial fulfillment of the require-**ment of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

(Prof. Dattatray Modani)
Internal Guide
Department of Computer Engineering
MCOE, PUNE

(Prof. Dr. Mrs. S. A. Itkar)
Head,
Department of Computer Engineering
MCOE, PUNE

(.......................)
External Examiner

(Dr. Mrs. K. R. Joshi)
Principal,
P.E.S Modern College of Engineering

Place: Pune.
Date: ...../...../2022

PROJECT APPROVAL SHEET

A Project Report Titled as

INTERACTIVE SYSTEM FOR PRODUCT PURCHASE THROUGH GESTURE AND VOICE

Is successfully completed by

Gaurav Zanpure	**Exam No: B150314352**
Bhavik Ransubhe	**Exam No: B150314320**
Adesh Oak	**Exam No: B150314299**
Rithvik Poojary	**Exam No: B150314323**

DEPARTMENT OF COMPUTER ENGINEERING

P.E.S MODERN COLLEGE OF ENGINEERING, PUNE-411005

SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE

ACADEMIC YEAR 2021-2022

Prof. Dattatray Modani	Prof. Dr. Mrs. S. A. Itkar
Internal Guide	H.O.D
Dept. of Computer Engg.	Dept. of Computer Engg.

Acknowledgment

It gives us great pleasure and satisfaction in presenting this project report on "**Interactive System For Product Purchase Through Gesture and Voice** ".

We thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of [Computer Dept] which helped us in successfully completing our project work. Also, We would like to extend our sincere esteems to all staff in laboratory for their timely support.

We have furthermore to thank Computer Department HOD Prof. Dr. Mrs. **S. A. Itkar and Guide Prof. Dattatray Modani to encourage us to go ahead and** for continuous guidance. We would also like to thank our project team members who showed immense patience and understanding throughout the project.

We would like to thank all those, who have directly or indirectly helped us for the completion of the work during this project.

Gaurav Zanpure

Bhavik Ransubhe

Adesh Oak

Rithvik Poojary

(B.E. Computer Engg.)

Abstract

E-commerce websites are becoming extremely prevalent, with a significant volume of individuals embracing this medium. This seems to be due to the enhanced accessibility and simplicity of use that these networks provide when it start shopping for necessities as well as other things from the convenience of your own home. This has enabled many handicapped people and others with accessibility concerns to successfully remain at home and purchase different products that they require. This has resulted in a plethora of e-commerce companies fighting for consumers' focus and attempting to grow their customer base in order to deliver accurate alternatives with simplicity. The integration of speech and gesture-based searching for items on e-commerce platforms is one example of something like this. The paradigm of utilizing Convolution Neural Network for the recognition of the gestures is one of the most difficult approaches that require extensive computations along with the region of interest estimation. The obtained results are indicate better performance of the system through the use of Root mean square Error evaluation.

Keywords: Convolutional Neural Network, Voice recognition and Gesture Recognition.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Overview

In current history, the E-Commerce sector has experienced tremendous expansion in internet marketplaces. This may be due to advancements in the online medium, which have resulted in a huge rise in the subscriber base, rendering most Web application implementations extremely cost. This really is critical for the growth of the information infrastructure, which has focused on improving the usefulness and accessibility of the product's consumers. The huge reduction in the cost of internet-enabled gadgets has accelerated the expansion, resulting in a massive boom with in E-Commerce sector.

There has already been a great deal of study into how to develop this platforms much further and supply consumers with novel and efficient factors that characterize it apart from its contemporaries. As a result, there has been an abundance of research towards increasing the customer satisfaction and providing unique characteristics that can engage a bigger percentage of subscribers. This empowers the website to expand as well as provide additional services, making it far easier for customers to purchase from the convenience of their own residences.

Humans have been seeking to comprehend human physical communication for millennia and have been quite effective in revealing intricacies of evolutionary cognition. Computers as well as other technologies have been used to investigate human feelings and make generalizations regarding human behavior because of complicated psychological interaction and detailed modes of behavior. To get a better comprehension of human emotions, scientists are studying hand signals, gestures, and expressions.

Historically, retailing was much more of a commercial operation, with shop performance reliant on the achievement of the business engagements. Selling products, managing expenses, and reaching profitability were the main priorities. With the importance of client brand recognition rising, the strategic management focuses from merely marketing to creating an engaging consumer experience.

A customer who walks into a retail establishment does not just want to purchase something. He wants to go on a voyage of contact and participation with a brand throughout several procuring rostrums. Customers' perceptions of how a buying experience should then have been shaped in large part by new tech. From the accessibility among several information sources collecting to the capacity to explore various possibilities for acquisition, compare characteristics, and find excellent prices, reading assessment information assists the contemporary potential in a variety of ways. Is it possible to persuade customers even more by providing a visual and tactile experience?

Customers' buying habits have shifted dramatically in the last few years. Purchasing must be viewed as a brand contact and involvement narrative, according to the company. This necessitates the development of a large number of interaction points. These observations may be earned at various times throughout the purchase process. The greater the degree of contact and participation that a business can achieve, the stronger the brand narrative will be. Testimonials are real-life events that help people connect with a business on an unconscious level. After this link is established, the consumer will return to the establishment not just to shop but also to maintain the connection.

Gesture recognition is utilized in a variety of industries, including automobiles, smart appliances, home appliances, sport, and defense, smartphones, and sign language interpretation. Among the most interesting jobs in the computer vision technique is recognizing these motions, which is tough for the machine to do. This presentation can help with the development of a beneficial element in e-commerce platforms that allows for fast browsing and engagement while purchasing things.

This is paired with the use of voice recognition, which are collected and utilized to explore and interface with the goods by the system. The technique for speech recognition is exceedingly intricate and difficult to understand. Voice command identification is used in multiple human integrates directly. A diverse variety of voice aided or voice command implementations have been commercialized as a result of advancements in speech identification. Examples of such applications are motorized wheelchair control, voice controlled motors, voice activated appliances, and voice controlled smartphone apps. Many papers have appeared in the literature on voice command recognition which have been evaluated effectively to understand its implementation and the gesture based product navigation for achieving our approach.

1.2 Motivation

E-commerce websites are becoming extremely prevalent, with a significant volume of individuals embracing this medium. This seems to be due to the enhanced accessibility and simplicity of use that these networks provide when it start shopping for necessities as well as other things from the convenience of your own home. This has enabled many handicapped people and others with accessibility concerns to successfully remain at home and purchase different products that they require. This has resulted in a plethora of e-commerce companies fighting for consumers' focus and attempting to grow their customer base in order to deliver accurate alternatives with simplicity. The integration of speech and gesture-based searching for items on e-commerce platforms is one example of something like this. The paradigm of utilizing image processing for the recognition of the gestures is one of the most difficult approaches that requires extensive computations. The implementation of voice based navigation through recognition has been studied in detail through the use of the related works.

1.3 Problem Definition and Objectives

To achieve effective realization of gesture and voice based interactive system through the efficient use of Convolutional Neural Networks and image processing methodologies.

1.4 Project Scope and Limitations

1.4.1 Project Scope

- The User Interface must be easy to use and should be easily understood by a layman. The interface must be clean and devoid of any artefacts and errors.
- The proposed system has been deployed with the use of completely free or open source software. The system does not employ any proprietary software or API's to achieve its goals.
- System is deployed in a standalone simulated environment.
- To effectively capture the image and identify the gesture.
- To allow voice navigation to the users easily.
- To implement the Convolutional Neural Networks accurately.

1.4.2 Limitations

- The database is moderate.
- Deployed in offline environment.
- Working in local host.
- The database is static.

1.5 Methodologies of Problem Solving

The performance metric of RMSE or Root Mean Square Error is one of the most effective performance metric to determine the error achieved between a set of continuous and correlated attributes. The attributes being selected for the evaluation of the proposed methodology are, hand gesture identified correctly and hand gesture identified incorrectly. The RMSE is calculated using the equation 1.1 given below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{i,l} - x_{2,i})}{n}} \qquad \text{———————}(1.1)$$

Where,

$\sum$ - Summation

(Zfi - Zoi)2 - Differences Squared for the hand gesture identified correctly and hand gesture identified incorrectly

N - Number of conducted Experiments

CHAPTER 2

LITERATURE SURVEY

This section of the literature survey eventually reveals some facts based on thoughtful analysis of many authors work as follows.

D. Ryumin et al. [1] presented a design for a smart robotic trolley for supermarkets with a touchscreen and a multimodal user interface that includes sign language and acoustic voice recognition, as well as a software framework for collecting sign language databases using the Kinect 2.0 device. The gathered corpus TheRusLan contains recordings of 13 natural Russian sign language signers. Each signer performed the same 164 sentences five times. The corpus contains a total of 10660 samples. TheRusLan will be utilized for more research and tests on a Russian sign language identification system that will be unsegregated into the smart retail trolley's conversation system.

Text mining was used by Preeti Mehra in this research to provide a unified understanding of terms in the retail business, which has been using gesture control technologies. A two-step technique including the extraction of keywords by text mining and then clustering these keywords into clusters was used to analyze data from retail customer evaluations [2]. Retailers have used gesture control technology in a variety of ways, including Kinect for Windows Retail Shopping, 'Gesture Control Screen,' 'Virtual Fitting Rooms,' 'Retail Interactive Touch Screens,' 'Use of Visual Mirrors,' and 'Augmented Reality Window Display'. Customer happiness is directly influenced by the 'Mechanism of Gesture' Control. The findings revealed the keywords that influence customer happiness as well as the features of this technology that influence their perception and help them make purchasing decisions.

Based on collaboration with a community in the Oku-tama district of Tokyo, Japan, Y. Shimizu et al. presented an interactive information support system for the rental bicycle industry. The authors used two distinct types of robot partners:

a concierge-type and a humanoid type with a tablet PC. Each robot companion has a distinct function to perform in conversation and suggestion [3]. First and foremost, the authors should address what robots should do and what people should do while providing customer support. Customers, for example, may find it difficult to respond to shop clerks' inquiries about their personal expenses for today. Customers can be asked by robot partners instead of shop assistants if shop staff do not wish to ask such a question.

D. Wu et al. introduced a unique Deep Dynamic Neural Network for continuous gesture detection on multimodal data that included image and depth data as well as skeletal characteristics In contrast to past state-of-the-art approaches, the authors do not rely on handmade features, which are time-consuming to develop, especially when done separately for each input modality [4]. Instead, deep neural networks are used to automatically extract useful information from data. Because the input data is multimodal, the presented model incorporates two separate feature learning methods: (1) Deep Belief Networks for skeletal feature processing and (2) 3D Convolutional Neural Networks for RGB-D data. In addition, the authors used an HMM to expand their feature learning model to include temporal relationships.

T. Du et al. suggested a deep learning framework for determining behavioral categorization. Deep learning offers strong feature extraction and categorization capabilities, and it has a high research value as a simulation of the biological neural network method framework. It is extremely important to recognize the human body motion, regardless of algorithm or application direction, using the advanced intelligence algorithm of deep learning and the information of human body movement acquired [5]. To perform gesture recognition, the RNN, LSTM, and GRU models are built. The experimental findings of the three models were compared. The results reveal that these approaches can identify hand motions in real-time, especially complicated gestures.

H. Long et al. introduce a deep learning-based gesture recognition framework that predicts the square anchor of the gesture and crop the sub-image from the original gesture using a two-stage recognition framework. The sub-image is recognized by the classifier network, and the category of the related gesture is determined [6]. The square anchor design is utilized to overcome the anchor deformation problem during scaling. Pruning and weight quantification techniques are used to further compress the model. The model approach may be applied in the embedded terminal while assuring the accuracy of gesture detection.

A computer vision technique is suggested by H. A. Jalab for operating a media player utilizing a neural network that identifies four hand gestures: play, stop, forward, and reverse. After capturing a frame from the webcam camera, skin segmentation in LAB color space was utilized to separate skin areas from background pixels. A fresh picture was produced that included the user's hand border. The form features of a hand gesture are described using a convex hull and corner detection [7]. A supervised back-propagation multi-layer feed-forward neural network was also utilized to classify user hand motions. The categorization was completed without the use of any special instruments, such as gloves or a marker. However, despite minor classification failures, the suggested system performed effectively in classifying the four users' hand gesture commands.

Faster R-CNN is used by H. Ruan et al. to recognize gestures based on pictures rebuilt with programmable metasurface. Depending on the gesture placements, the authors enhance the imaging of these gesture areas and achieve high accuracy recognition of 10 types of gestures using CNN and high-resolution pictures. Gesture detection and identification may be quite useful in assisting people's communication in various scenarios [8].

S. Hussain et al. suggest a transfer learning and vision-based hand gesture recognition technique for unidirectional dynamic motions, the approach was made more robust by omitting skin color segmentation, blob detection, skin area cropping, and centroid extraction. The pre-trained model in the presented work is VGG16, a CNN architecture. It has 13 convolution layers followed by three completely linked layers. A convolutional neural network (CNN) is a sort of feed-forward neural network whose connection structure is inspired by the arrangement of the animal visual brain [9]. Because the authors need to distinguish eleven different hand forms, CNN is trained as a classifier using the transfer learning approach.

With an accuracy of 93.09 percent, the prototype was successfully tested on seven different volunteers in various backdrops and lighting situations.

The convolutional neural network framework in deep learning was utilized by Y. Gu et al. [10] to create an online teaching gesture identification model that can be used to recognize five distinct types of gestures that teachers will employ in online teaching settings. The model's recognition efficiency and accuracy are quite good. The model may be used to investigate the impact and quality of gestures in teaching, as well as the enhancement of educational robots' performance in human-computer interaction. However, at this time, the model is unable to distinguish between symbolic and metaphorical gestures that must be identified depending on the language utilized by the teachers or the actual teaching environment.

Y. Zhang provides a conceptual model to investigate the impact of traditional and virtual communities on customers' choice to purchase a product online. It makes theoretical as well as practical contributions [11]. It integrates conventional and virtual communities and presents a conceptual model to examine the aspects that affect a buyer's need recognition in online buying more thoroughly by evaluating the structure of virtual communities and their influence on FNR. In terms of practical contribution, it suggests that the firm should consider word of mouth in both the traditional and virtual communities. Both communities are critical for potential customers to identify their requirements and create buy intents.

X. Hu. et al. employ a large data analysis approach, the deep forest algorithm, to develop a prediction model using real consumer online buying behavior data, to predict customer purchase behavior in the context of online purchasing [12]. The outcomes of the online purchasing behavior prediction model depending on the deep forest algorithm are superior to other models in some circumstances, according to actual evidence. Because multiple models may be cascaded with deep forests, such as substituting the cascaded forest model with a linear regression model, the model in this research has the potential to enhance the classification prediction impact even further.

H. Cheng presents a research model that depends on social cognition theory to investigate the links between trust, perceived website complexity, and online buying behavior. Furthermore, earlier research has found that the online shop atmosphere is a major component affecting online purchases. However, perceived website complexity is given less weight. Therefore, this research considers perceived website complexity to be an environmental element and investigates the association between perceived website complexity and online purchasing behavior [13].

CHAPTER 3

SOFTWARE REQUIREMENTS AND SPECIFICATION

3.1 Introduction

This Software Requirements Specification provides a complete description of all the functions and constraints of the **"Interactive System For Product Purchase Through Gesture and Voice ".** The document describes the issues related to the system and what actions are to be performed by the development team in order to come up with a better solution.

The notion of gesture recognition, as it is often known, entails the interpretation of human motions using mathematical algorithms. The study entails interpreting emotions through movements, most of which originate on the face or in the hands. These motions are used to decipher human body language and infer human ideas and orders. Gesture recognition is employed in a variety of industries, including Automative, Home Automation Consumer Electronics, Gaming, Defense, Mobiles, sign language translation, and so on.

3.1.1 Project Scope

The purpose of this SRS document is to provide a detailed overview of our software product "**Interactive System For Product Purchase Through Gesture and Voice** ", its parameters and goals. This document describes the project's target audience and its user interface, hardware and software requirements. It defines how our client, team and audience see the product and its functionality.

The scope of this project includes project developer assisted by project guide. The scope thus far has been the completion of the basic interfaces that will be used to build the system. The proposed model should detect successfully with sober user interface and settings. The constraints felt thus far by the developer have only been our weekly story cards, the end-to-end side of the interface, and time to time brushing on methodology of implementation which schedule the completion of the project.

The major scope of this project is as follows

1. The User Interface must be easy to use and should be easily understood by a layman. The interface must be clean and devoid of any artefacts and errors.

2. The proposed system has been deployed with the use of completely free or open source software. The system does not employ any proprietary software or API's to achieve its goals.

3. System is deployed in a standalone simulated environment.

4. To effectively capture the image and identify the gesture.

5. To allow voice navigation to the users easily.

6. To implement the Convolutional Neural Networks accurately.

3.1.2 User Classes and Characteristics

This project is for the user who are interested in voice and gesture based shopping. This is having following characteristics:

- Register.
- Login.
- Update.

3.1.3 Assumption and Dependencies

Following are the assumptions:

- System is deployed in single machine.
- Database is given required permissions.

Dependencies can be:

- System speed.
- IDE behavior.

3.2 Functional Requirements

3.2.1 Preprocessing

3.2.1.1 Description and Priority

- Input-Image dataset
- Priority –High

3.2.1.2 Stimulus/Response Sequences

- Stimulus: Image Scaling.
- Response sequence: Image sharing and restoration

3.2.1.3 Functional Requirements

- Dataset list formation

3.2.2 Image Normalization

3.2.2.1 Description and Priority

- Input-Images
- Priority –High

3.2.2.2 Stimulus/Response Sequences

- Stimulus: Colour Model.
- Response sequence: Model features

3.2.2.3 Functional Requirements

- Region Estimation

3.2.3 CNN

3.2.3.1 Description and Priority

- Input-Skin Images
- Priority –High

3.2.3.2 Stimulus/Response Sequences

- Stimulus: activation Function.
- Response sequence: Kernel Formation.

3.2.3.3 Functional Requirements

- Output Layer prediction Score.

3.2.4 Decision Making

3.2.4.1 Description and Priority

- Input-Voice
- Priority –High

3.2.4.2 Stimulus/Response Sequences

- Stimulus: Word identification
- Response sequence: IF-then Rules

3.2.4.3 Functional Requirements

- Automatic Interface Handling for Shopping

3.3 External Interface Requirements

3.3.1 User Interfaces

Our system interacts with user on the following occasions:

1. By the user while viewing the results.

3.3.2 Hardware Interfaces

Our system interacts with hardware components on the following occasions:

- Our system interacts with secondary memory device while performing database operations.

3.3.3 Software Interfaces

Our system interacts with the Mysql Database server while performing database operations.

3.3.4 Communication Interfaces

Our system's different modules are communicating with one another on the following scenarios:

1. From capture voice and gesture module to voice recognition module.
2. From voice recognition module to display products module.
3. From display products module to gesture recognition module.
4. From Gesture dataset module to preprocessing module.
5. From preprocessing Module to CNN module.
6. From CNN module to gesture recognition Module.

3.4 Nonfunctional Requirements

3.4.1 Performance Requirements

- **High Speed:** System should process voice and gesture recognition in good speed. Then system must wait for process completion.
- **Accuracy:** System should correctly execute process, display the result accurately. System output should be in user required format. That means all the intermediate steps need to be display properly.

3.4.2 Safety Requirements

The system need to be properly converted into executable file format so that the source code of the system will preserved from any kind of intrusions and database should provide proper credentials.

3.4.3 Security Requirements

Database should be given proper username and passwords where protocols and other data is being stored for the system usage.

3.4.4 Software Quality Attributes

- Number of functionality provided must be properly working independently.
- System should be reliable.
- All logins are secured.
- System provide accurate result of CNN module.
- Survivability of the system should be good
- System performs each function accurately.
- System is extensible.

3.5 System Requirements

3.5.1 Database Requirements

MYSQL 5.5

3.5.2 Software Requirements(Platform Choice)

1. Platform: JAVA and PYTHON
2. Technology : JDK 1.8 and Python 3.8
3. IDE: Netbeans 8.2 and Spyder 4.2

3.5.3 Hardware Requirements

1 systems of following minimum configuration

Sr. No.	Parameter	Minimum Requirement	Justification
1	Processor	2.2 GHz	For Fast Processing
2	Hard Disk	201 GB	For Fast Processing
3	RAM	2 GB	For Fast Processing
4	Monitor, Keyboard and UPS	1	None
5	Web Cam	1	Logitech

Table 3.1: Hardware Requirements

3.6 System Implementation Plan

Figure 3.1: Timeline Chart

CHAPTER 4

SYSTEM DESIGN

4.1 System Architecture

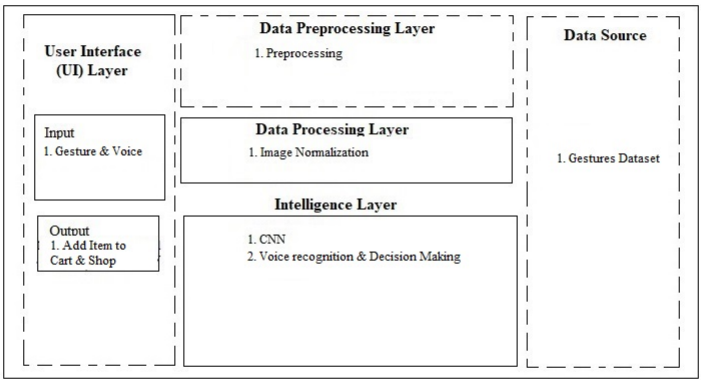


Figure 4.1: System Architecture

4.2 System Design

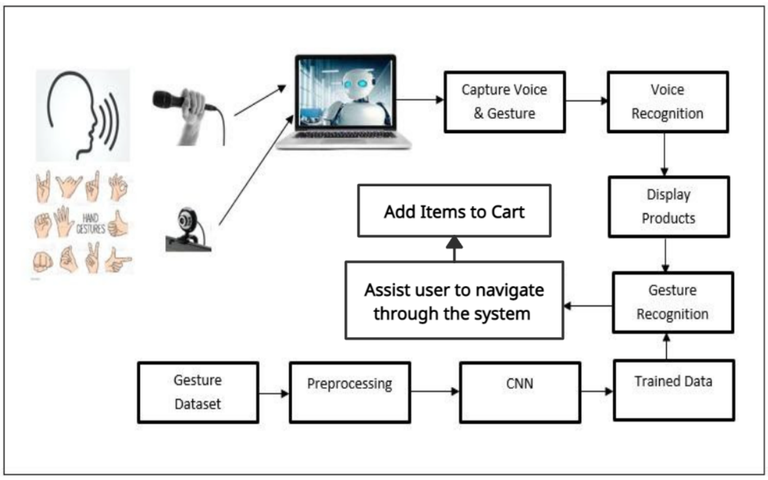


Figure 4.2: System Design

4.3 Mathematical Model

(A) Set Theory

1. S= {} be as system for Shopping System through Gesture & Voice Assistance

2. Identify Input as S=$\{G_1, G_2, G_3.....G_n \& V_1, V_2, V_3.....V_n\}$

   Where G = Gesture & V= Voice

   S=$\{G, V\}$

3. Identify A_{ITC} as Output i.e. Add Item to Cart

   S= $\{G, V, A_{ITC}\}$

4. Identify Process P

   S= $\{G, V, P, A_{ITC}\}$

   P= $\{P_R, I_N, C_{NN}, D_M\}$

   Where P_R=Preprocessing

   I_N=Image Normalization

   C_{NN} = Convolutional Neural Network

   D_M = Decision Making

   So complete system for Voice and Gesture Based Shopping System can be given as

5. S=$\{G, V, P_R, I_N, C_{NN}, D_M, A_{ITC}\}$

(B) Set Description

1. **Preprocessing**

   Set P_R:

   P_{R0} = Image Scaling

   P_{R1}=Image Sharing

   P_{R2}=Image restoration

   P_{R3}= Dataset list formation

2. **Image Normalization** Set I_N:

   I_{N0} =Pixel Position

   I_{N1}=Color Model

   I_{N2}=Model Features

   I_{N3}= Region Estimation

3. **Convolutional Neural Network**

   Set C_{NN}:

   C_{NN1}= ROI Region

   C_{NN2}= First Layer Convolution

   C_{NN3}= Fully Connected layer

   C_{NN4}= Convolution Rate

4. **Decision Making**

   Set D_M:

   D_{M0}=Voice access

   D_{M1}=Word Identification

   D_{M2}=IF-then Rules

   D_{M3}= Automatic Interface Handling for Shopping

(C) Representation of Sets and its operation

1. **Union Representation**

A. Set $P_R = \{P_{R0}, P_{R1}, P_{R2}, P_{R3}\}$ Set $I_N = \{I_{N0}, I_{N1}, I_{N2}, I_{N3}\}$

   Set $(P_R U I_N) = \{P_{R0}, P_{R1}, P_{R2}, P_{R3} U I_{N0}, I_{N1}, I_{N2}, I_{N3}\}$

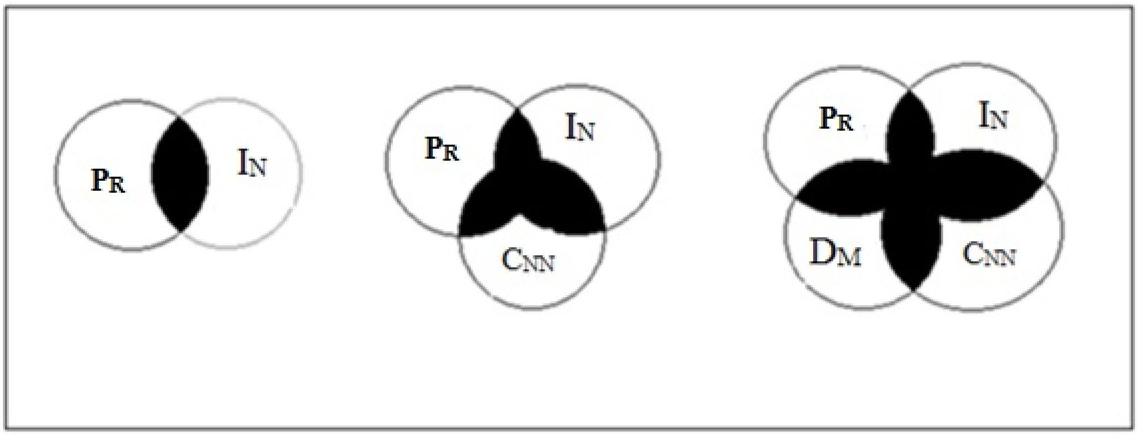
B. Set $C_{NN} = \{C_{NN0}, C_{NN1}, C_{NN2}, C_{NN3}\}$

   Set $(P_R U I_N) = \{P_{R0}, P_{R1}, P_{R2}, P_{R3} U I_{N0}, I_{N1}, I_{N2}, I_{N3} U C_{NN0}, C_{NN1}, C_{NN2}, C_{NN3}\}$

B. Set $D_M = \{D_{M0}, D_{M1}, D_{M2}, D_{M3}\}$

   Set $(P_R U I_N) = \{P_{R0}, P_{R1}, P_{R2}, P_{R3} U I_{N0}, I_{N1}, I_{N2}, I_{N3} U C_{NN0}, C_{NN1}, C_{NN2}, C_{NN3}$
   $U D_{M0}, D_{M1}, D_{M2}, D_{M3}\}$

2. **Venn Diagram**



4.4 Data Flow Diagrams

4.4.1 DFD Level 0

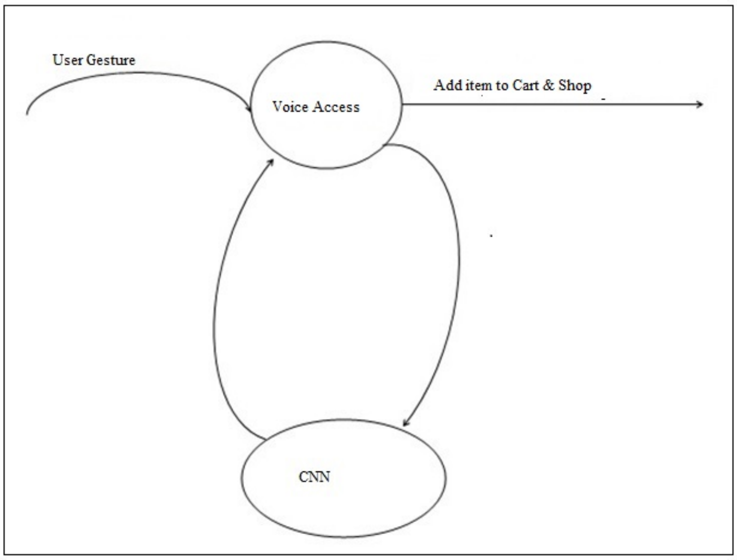


Figure 4.3: DFD Level 0 Diagram

4.4.2 DFD Level 1

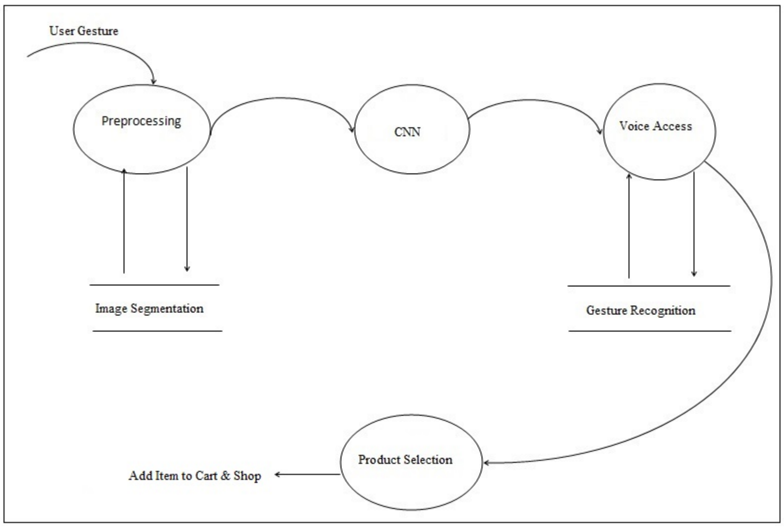


Figure 4.4: DFD Level 1 Diagram

4.4.3 DFD Level 2

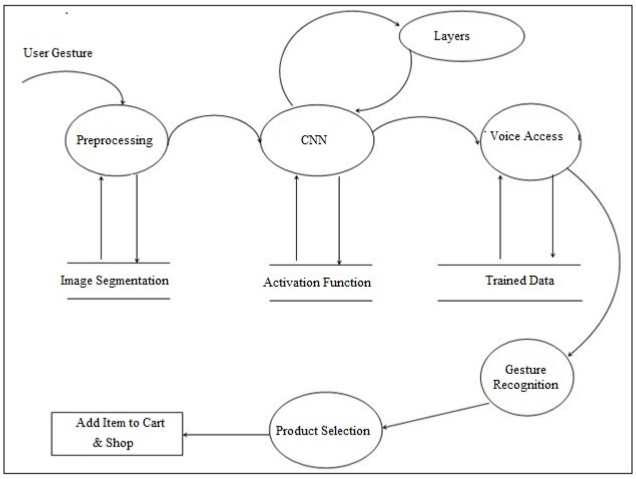


Figure 4.5: DFD Level 2 Diagram

4.5 UML Diagrams

4.5.1 Use Case Diagram

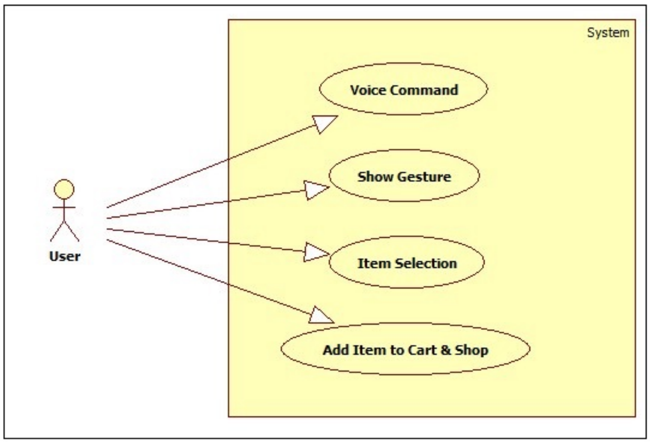


Figure 4.6: Use Case Diagram

4.5.2 Activity Diagram

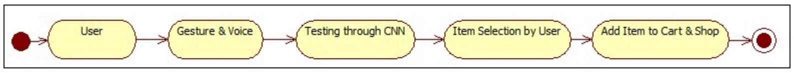


Figure 4.7: Activity Diagram

4.5.3 Sequence Diagram

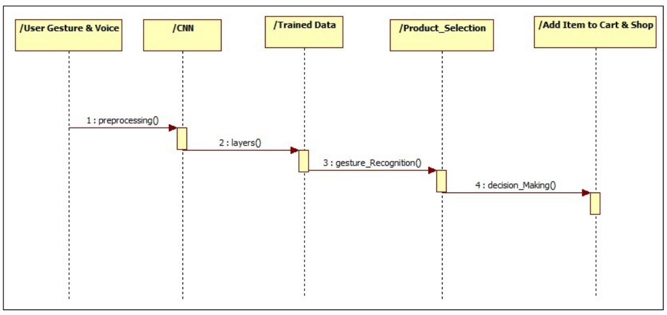


Figure 4.8: Sequence Diagram

4.5.4 Component Diagram

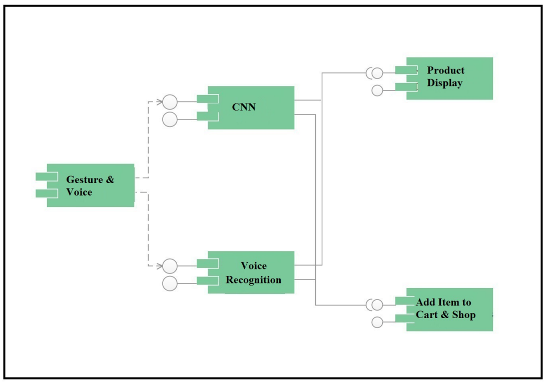


Figure 4.9: Component Diagram

4.5.5 Deployment Diagram

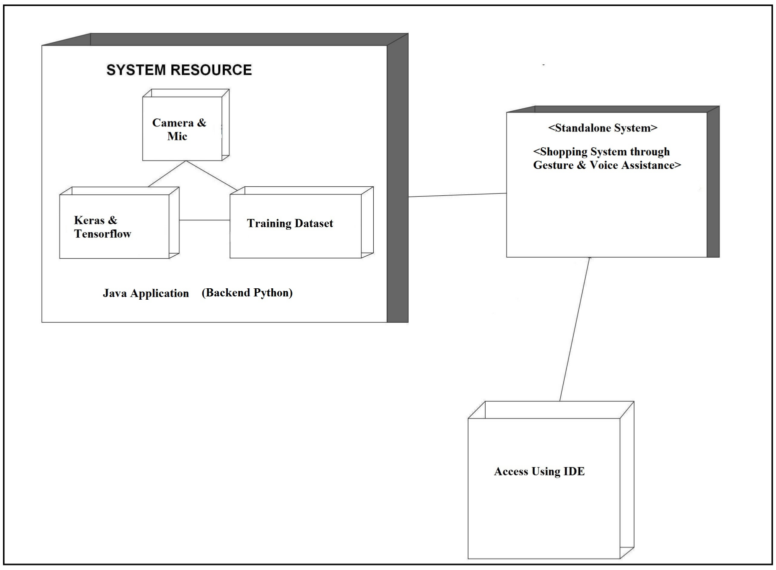


Figure 4.10: Deployment Diagram

4.5.6 Package Diagram

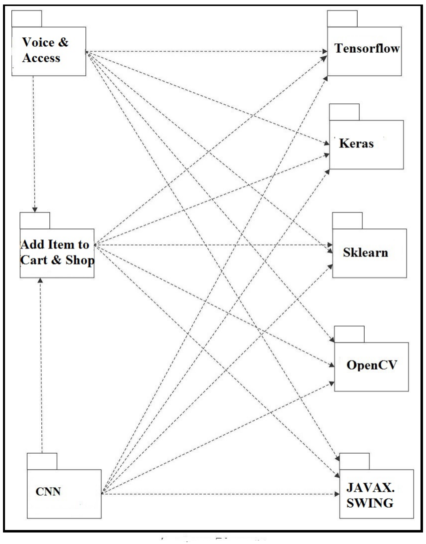


Figure 4.11: Package Diagram

4.5.7 State Transition Diagram

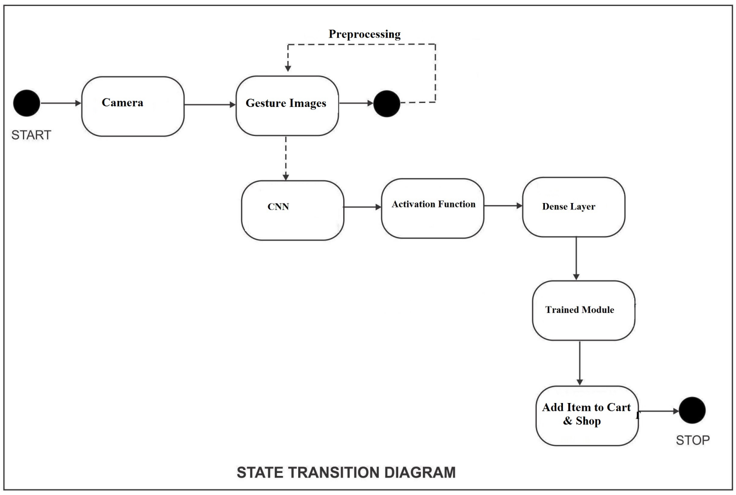


Figure 4.12: State Transition Diagram

CHAPTER 5

PROJECT PLAN

5.1 Project Estimation

5.1.1 Time Estimation of KLOC

The no of lines required for implementation of various modules can be estimated as follows:

Function	Estimated KLOC
Preprocessing	0.226
Image Segmentation	0.230
CNN	1.135
Decision Making	1.267
Gesture Recognition	0.222

Table 5.1: Estimation in KLOC

Thus the total number of lines required is approximately 3.15 KLOC Efforts:

E = 3.15 * (3.08) ^ 1.02 (According to COCOMO Model)

E = 3.15 * (3.15)

E = 9.9225

Development Time for Implementation and Testing

D = 9.9225 Months is the development Time needed for Project.

D=Development Time for Project.

5.1.2 Cost Estimate

This cost can be calculated with the standard pay roll in India assigned to the fresher software developer according to industry standard. So let we take some amount as X per hour for development.

So cost can be calculated as below.

Number of minimum hours per developer for a month = 10

So Total number of hours = 10 *10*4 $\Rightarrow$ 400 Hours

So, Total cost can be said as 400XRs.

5.2 Risk Management w.r.t. NP Hard analysis

5.2.1 Risk Identification

Answers to following questionnaire revealed some risks.

1. Have top software and customer managers formally committed to support the project?

   Answer : Yes

2. Are end-users enthusiastically committed to the project and the system/product to be built?

   Answer : Yes

3. Are requirements fully understood by the software engineering team and its customers?

   Answer : Yes

4. Have customers been involved fully in the definition of requirements?

   Answer : Yes

5. Do end-users have realistic expectations?

   Answer : Yes they can cross verify got result for gesture recognition.

6. Does the software engineering team have the right mix of skills?

   Answer : Yes

7. Are project requirements stable?

   Answer : No , they can be change depend on the usage. As of now we are hosting in LAN so they are stable for LAN

8. Is the number of people on the project team adequate to do the job?

   Answer : Yes

9. Do all customer/user constituencies agree on the importance of the project and on the requirements for the system product to be built?

   Answer : Yes

5.2.2 Risk Analysis

ID	Risk Description	Probability	Impact on Schedule	Impact on Quality	Impact on Overall
1	Image Size	Low	Low	Low	Low
2	Improper Input	Medium	Medium	Medium	Medium

Table 5.2: Risk Table

Probability	Value	Description
High	Probability of occurrence is	>75%
Medium	Probability of occurrence is	26-75%
Low	Probability of occurrence is	<25%

Table 5.3: Risk Probability Definition

5.2.3 Overview of Risk Mitigation, Monitoring, Management

Following are the details for each risk:

5.3 Project Schedule

5.3.1 Project task set

Major Tasks in the Project stages are:

Impact	Value	Description
Very High	>10%	Schedule impact or Unacceptable quality
High	5-10%	Schedule impact or Some parts of the project have low quality
Medium	<5%	Schedule impact or Barely noticeable degradation in quality Low Impact on schedule or Quality can be incorporated

Table 5.4: Risk Impact Definition

Risk ID	1
Risk Description	Image Size
Category	User Manual
Source	SRS
Probability	LOW
Impact	High
Response	System Can Stop Responding
Strategy	Can alert while testing itself
Risk Status	Not Yet Occurred

Table 5.5: Risk Monitoring

Risk ID	2
Risk Description	Improper Input
Category	User Manual
Source	SRS
Probability	Medium
Impact	High
Response	System Can Stop Responding
Strategy	Can validate and Rise Alert
Risk Status	Not Yet Occurred

Table 5.6: Risk Management

Sr. No.	Tasks	Description
1	GUI Creation	Proper GUI Designing using Swing Frame Work
2	Validation	Maintaining Time slice
3	Testing	Testing the Module
4	Analysis	Result Analysis for RMSE Parameter

Table 5.7: Major Task

5.4 Team Organization

The manner in which staff is organized and the mechanisms for reporting are noted.

Sr. No.	Name	Organization
1	Gaurav Zanpure	Requirement Gathering, Developing, Testing and Reporting
2	Bhavik Ransubhe	
3	Adesh Oak	
4	Rithvik Poojary	

Table 5.8: Team Organization

5.4.1 Team structure

The team structure for the project is identified. Roles are defined.

Guide Name	Team Members
Prof. Namrata Adhao	Gaurav Zanpure
	Bhavik Ransubhe
	Adesh Oak
	Rithvik Poojary

Table 5.9: Team Structure

CHAPTER 6

PROJECT IMPLEMENTATION

6.1 Introduction

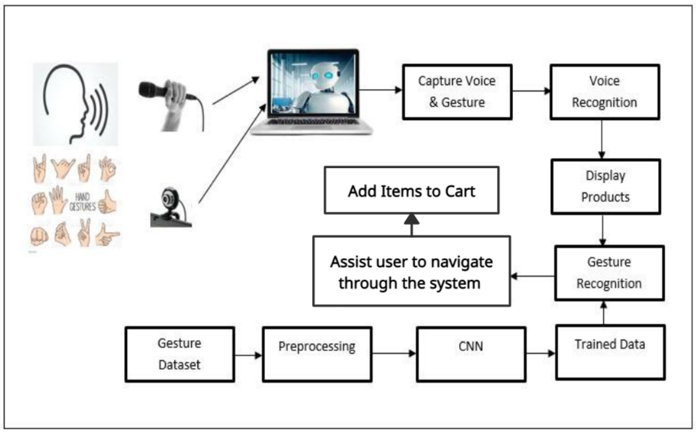


Figure 6.1: System Overview

The suggested approach for gesture-based shopping using hand gesture identification is detailed in the following step-by-step process. The methodology's system overview is depicted in Figure 6.1.

Step 1: Preprocessing – This is the initial phase in the process, in which the hand gesture images are taken using the OpenCv package. The cv2 library's VideoCapture capability is used to take photos of the specific hand gesture. There are five main forms of hand gestures: up, down, left, right, and buy. The YCbCr color model is used to identify the skin of the hand, and that particular region is clipped out. After that, the grayscale transformation is applied to the clipped hand gesture image. This produces a grayscale cropped picture, which is then downsized to 48x48 pixels and saved in a folder dedicated to that particular gesture. This process is performed iteratively for each of the 5 gestures for the purpose of achieving the input dataset.

Step 2: Image Segmentation – The input dataset, which comprises of acquired images that must be used for training, is made available in the previous stage. The training generator and the validation generator are in charge of this task. The training generator starts with a target picture size of 48x48 pixels, a batch size of 64, and a color mode of grayscale with categorical class mode.

The validation generator is also created with comparable qualities, such as a target picture size of 48x48 pixels, a batch size of 64, and a color model of grayscale with a categorical class mode.

Step 3: Convolution Neural Network – This is the most important part of the suggested method, since it is in charge of detecting and identifying hand gestures. In the convolutional Neural Network module, the original image is utilized as an input. This model is trained using the input photographs that were acquired, preprocessed, and segregated in the previous rounds of the technique.

The input dataset is made up of the training and testing picture folders. Each folder is then separated into separate directories for each hand gesture and its associated photos. As component of the training phase, these photos are input into the CNN model. The photos should first be resized to 48×48 pixels in width and height. The model is trained on these images for 500 epochs with a batch size of 64 and a dense of 5 because we are considering 5 gestures in our presentation. The TensorFlow and Keras libraries are used in the python environment to facilitate the different components of the CNN model. The architecture may be seen in diagram 2 below.

After that, the CNN model created with this architecture is run for 500 epochs to produce a trained model file with the extension.h5.

Step 4: Decision Making – After achieving the trained model through CNN, the technique may now be evaluated for hand gesture detection for shopping. The camera is activated with the help of the OpenCV platform in order to crop the hand gestures. This cropped mage is transformed to grayscale and scaled to work with the.h5 file's contents. This procedure produces the matching hand gesture, which is then sorted with all of the matched gestures at the time. If the acquired count exceeds a certain threshold, the gesture is considered to have been detected and used for shopping purposes.

Layer	Activation
CONV 2D 32 X 3 X 3	Relu
CONV 2D 64 X 3 X 3	Relu
MaxPooling2D 2 X 2	
Dropout 0.25	
CONV 2D 128 X 3 X 3	Relu
MaxPooling2D 2 X 2	
CONV 2D 128 X 3 X 3	Relu
MaxPooling2D 2 X 2	
Dropout 0.25	
Flatten	
Dense 1024	Relu
Dropout 0.25	
Dense 5	Softmax
Adam Optimizer	

Figure 6.2: CNN network Architecture

6.2 Tools and Technologies Used

The proposed methodology for achieving gesture based system through hand gesture recognition has been deployed on both the python programming language using the Spyder IDE and Java programming language using the NetBeans IDE. The approach utilizes the OpenCV, TensorFlow, and Keras, libraries to achieve the desired goals. The presented technique has been deployed on a computer consisting of 8 GB of RAM, 1 TB of Storage powered by an Intel Core i5 as the CPU.

CHAPTER 7

SOFTWARE TESTING

7.1 Type of Testing Used

7.1.1 Performance Testing

The performance of the system is measured by the Root mean square for the given input Gesture image dataset.

7.1.2 System Testing

Checked the steadiness of the space provided by the system for the maximum number of images. This is done by feeding the maximum number of images to learn and so that set the threshold by the Virtual machine option of the Spyder IDE.

7.1.3 Recovery Testing

Our system can be recovered in span of 2 hour after crashing. Within two hour we can install all the Software and deploy our system to work as in the past.

7.1.4 Security Testing

7.1.4.1 Stress Testing

The System is well equipped to stand against the breakdown point of maximum number of datasets decided by the IDE's virtual machine settings, beyond that the memory overflow exception may arise.

7.1.4.2 Unit Testing

All the modules are independently handled developed and ran to get proper output and finally they are integrated to get the whole output.

7.1.4.3 Black Box Testing

Compatibility analysis is done by passing the output of one module to another and checking for estimated output of Gesture Detection.

7.1.5 Integration Testing

When all the individual modules are integrated into our system, that yields a proper system of Gesture detection system that is cross examined for the desired output.

7.2 Test Cases and Test Results

7.2.1 Test Cases for Proposed System

Test Cases for User

ID	TEST CASE	INPUT	PASS CRITERIA
U_TRN	User Training the model	Dataset	Model is successfully trained and .h5 file is being created.
AD_LOGIN	Admin Login	Admin Username and Login	Successful login when valid credentials entered.
AD_EDT_PRFLE	Admin Edits the Profile	New Password	Successfully Credentials are updated.
AD_NEW_DET	Admin Adding the Item Details	Item Details	All item details are successfully Stored.
AD_NEW_EDT	Admin Adding the Item Details	Item Details	All item details are successfully Edited.
AD_GST_ENG	Admin Starts the Gesture Engine	Invoke Gesture Button	Gesture based shopping Engine Activates.

Table 7.1: Test Cases For User

CHAPTER 8

RESULTS

8.1 Outcomes

The proposed methodology for achieving gesture based system through hand gesture recognition has been deployed on both the python programming language using the Spyder IDE and Java programming language using the NetBeans IDE. The approach utilizes the OpenCV, TensorFlow, and Keras, libraries to achieve the desired goals. The presented technique has been deployed on a computer consisting of 8 GB of RAM, 1 TB of Storage powered by an Intel Core i5 as the CPU.

For the evaluation purposes the proposed model is trained for 500 epochs on 5 gestures and the outcomes for the same need to be evaluated for the purpose of achieving effective performance of the approach.

The accuracy of the hand gesture recognition module needs to be assessed in order to determine the performance of the proposed approach. The accuracy of the approach can be calculated as a metric of error, as lower the error, greater is the accuracy. The evaluation of error can be effectively performed using RMSE metric.

The performance metric of RMSE or Root Mean Square Error is one of the most effective performance metric to determine the error achieved between a set of continuous and correlated attributes. The attributes being selected for the evaluation of the proposed methodology are, hand gesture identified correctly and hand gesture identified incorrectly. The RMSE is calculated using the equation 8.1 given below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{i,l} - x_{2,i})}{n}} \qquad \text{——————}(8.1)$$

Where,

$\sum$ - Summation

(Zfi - Zoi)2 - Differences Squared for the hand gesture identified correctly and hand gesture identified incorrectly

N - Number of conducted Experiments

The RMSE values are computed for a number of iterations of hand gesture recognition performed through this proposed approach. Each of the 5 hand gestures are tested for the recognition 10 times. Each of the times the recognition output of the proposed approach is recorded. The outcomes are then utilized for the purpose of RMSE evaluation. These values of RMSE are rigorously calculated with the outcomes stipulated in the table 8.1 given below.

Gesture	Number of iterations	Correctly identified hand gesture	Incorrectly identified hand gesture	MSE
Top	10	9	1	1
Bottom	10	8	2	4
Left	10	9	1	1
Right	10	7	3	9
Bye	10	10	0	0

Table 8.1: RMSE outcomes for 5 hand gesture recognition

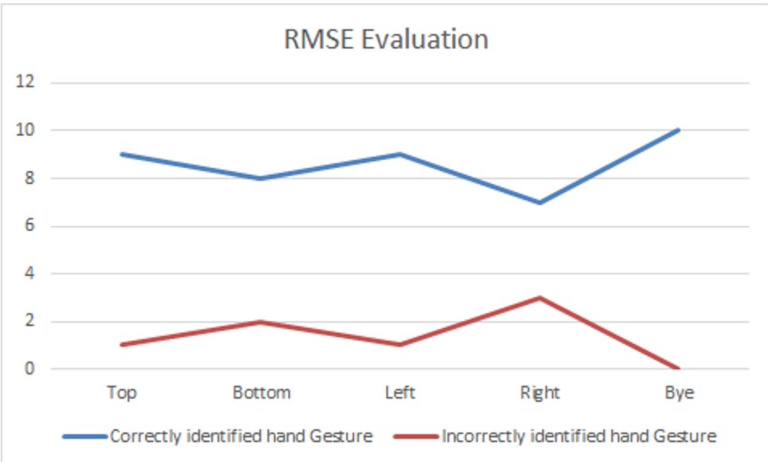


Figure 8.1: Line Graph for RMSE outcomes for 5 hand gesture recognition

The results attained for the recognition performance and RMSE values in the table 8.1 given above are being utilized for the purpose of achieving the line graph in the figure 8.1 given above. The graph and table above illustrate the intended methodology for hand gesture recognition achieving an incredibly low error rate. The higher recognition accuracy may be attributed to the suggested approach's use of deep learning via CNN, which enhances recognition accuracy substantially. The hand gesture recognition error has an RMSE of 3.8, which is a really good outcome in the very first try of the proposed work.

8.2 Screen Shots

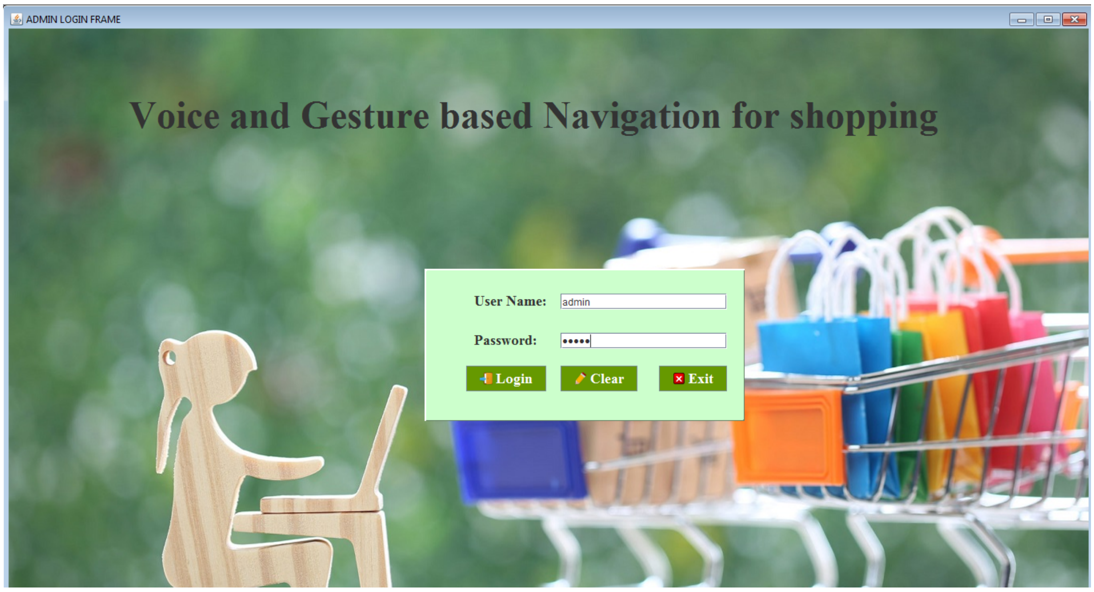


Figure 8.2: Login Frame

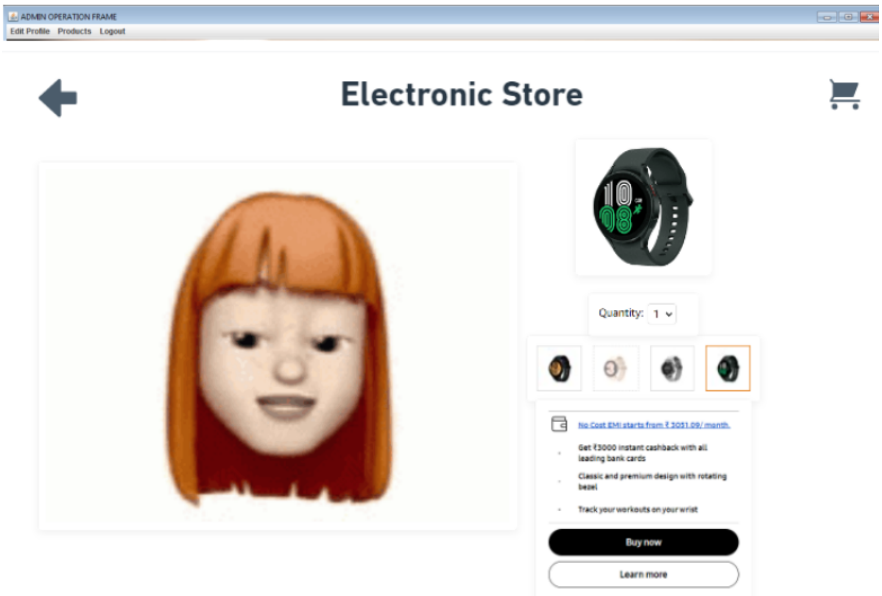


Figure 8.3: Operation Frame

Figure 8.4: Update Frame

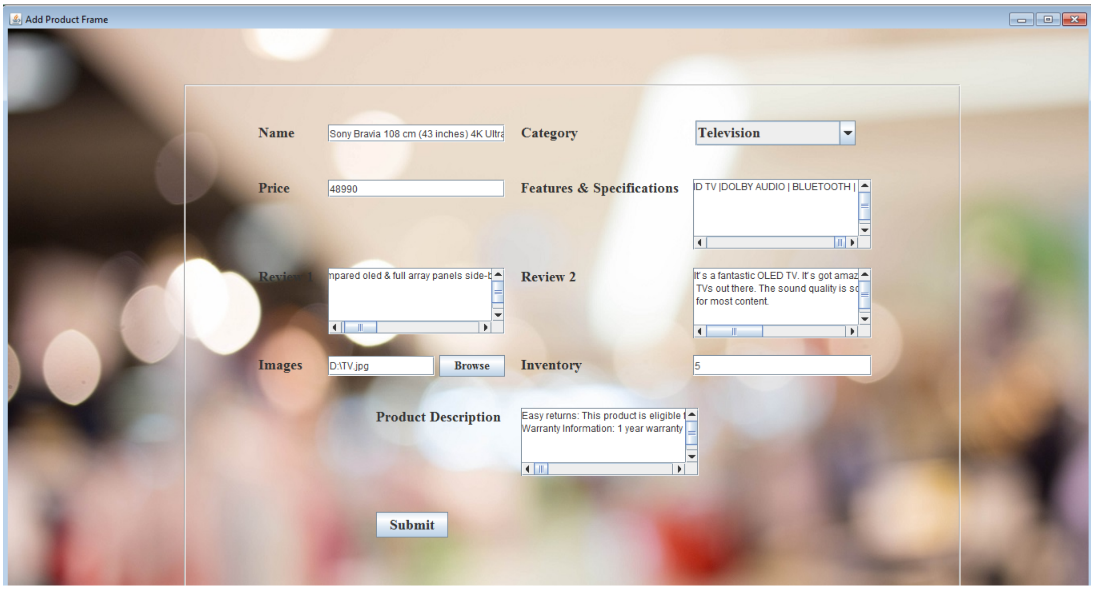


Figure 8.5: Add Product Frame

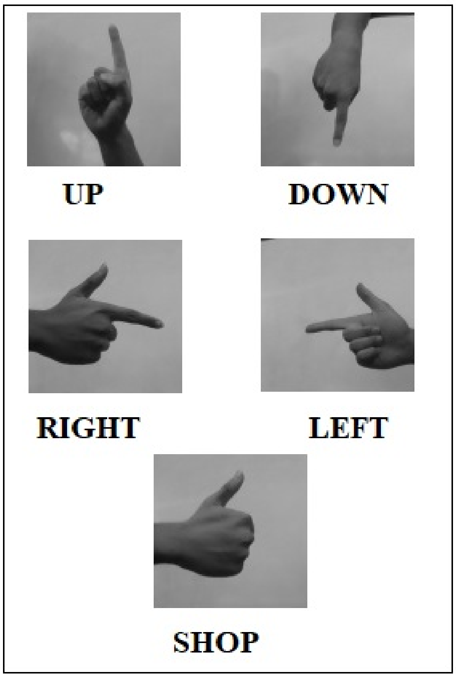


Figure 8.6: Outcome

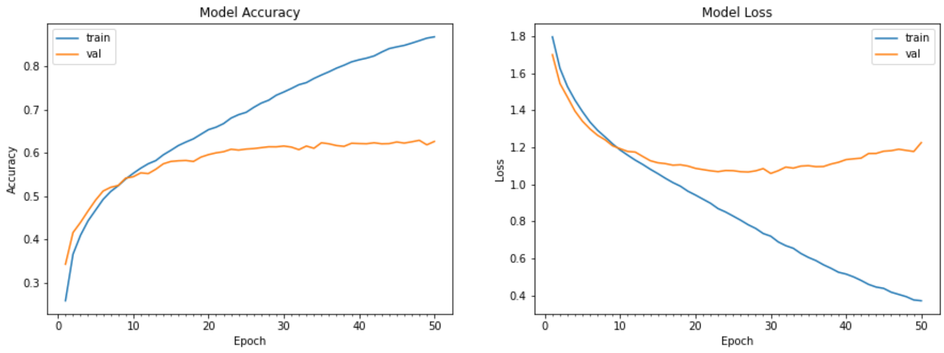


Figure 8.7: Accuracy and Loss Graph

CHAPTER 9

CONCLUSIONS

9.1 Conclusion

With the rapid development of computer technology, efficient human-computer interaction has now become an indispensable part of people's daily life. The most commonly used human-computer interaction mode is to rely on simple mechanical devices, such as mouse, keyboard, touch screen, etc. Through these typical controllers, it is difficult to achieve an immersive control experience. For example, handles or data gloves are usually used for interaction in a virtual reality environment, and bulky devices seriously affect the user's immersive experience. Therefore, the development of a more immersive human-computer interaction method has received widespread attention, and it is worth noting that the human computer interaction method through gesture recognition has been recognized by people. The hand is a complex deformable body with multiple degrees of freedom. The addition of voice based interaction also improves the overall user experience by improving the effective interaction between the user and the product. Hence, Convolution neural network is used to handle gesture and multi- threaded environment allows handling the graphical user interface efficiently.

9.2 Future Scope

- System can be enhance to work as independent API.
- System can enhance to work on all sectors.
- System can enhance to work on more number of protocols.

9.3 Application

- The approach can be applied on E-commerce websites to improve customer satisfaction.
- The system can be implemented in existing online portals and websites to improve the sales for the seller.

REFERENCES

[1] D. Ryumin, D. Ivanko, A. Axyonov, I. Kagirov, A. Karpov and M. Zelezny, "Human-Robot Interaction with Smart Shopping Trolley Using Sign Language: Data Collection," 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2019, pp. 949-954, DOI: 10.1109/PERCOMW.2019.8730886.

[2] Preeti Mehra and Balpreet Kaur, "Gesture Recognition: towards Making Future Retail Buying Experience Stimulating," International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-6, March 2020.

[3] Y. Shimizu, S. Yoshida, J. Shimazaki, and N. Kubota, "An interactive support system for activating shopping streets using robot partners in informationally structured space," 2013 IEEE Workshop on Advanced Robotics and its Social Impacts, 2013, pp. 70-75, DOI: 10.1109/ARSO.2013.6705508.

[4] D. Wu et al., "Deep Dynamic Neural Networks for Multimodal Gesture Segmentation and Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 8, pp. 1583-1597, 1 Aug. 2016, DOI: 10.1109/TPAMI.2016.2537340.

[5] T. Du, X. Ren, and H. Li, "Gesture recognition method based on deep learning," 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), 2018, pp. 782-787, DOI: 10.1109/YAC.2018.8406477.

[6] H. Long, M. Liu, and M. Li, "Real-time gesture recognition in the complex background based on deep learning," 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2020, pp. 2137-2140, DOI: 10.1109/ITAIC49862.2020.9339060.

[7] H. A. Jalab and H. K. Omer, "Human-computer interface using hand gesture recognition based on neural network," 2015 5th National Symposium on

Information Technology: Towards New Smart World (NSITNSW), 2015, pp. 1-6, DOI: 10.1109/NSITNSW.2015.7176391.

[8] H. Ruan, M. Wei, H. Zhao, H. Li, and L. Li, "Programmable Metasurface Based Microwave Gesture Detection and Recognition Using Deep Learning," 2020 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), 2020, pp. 1-4, DOI: 10.1109/NEMO49486.2020.9343444.

[9] S. Hussain, R. Saxena, X. Han, J. A. Khan, and H. Shin, "Hand gesture recognition using deep learning," 2017 International SoC Design Conference (ISOCC), 2017, pp. 48-49, DOI: 10.1109/ISOCC.2017.8368821.

[10] Y. Gu, J. Hu, Y. Zhou, and L. Lu, "Online Teaching Gestures Recognition Model Based on Deep Learning," 2020 International Conference on Networking and Network Applications (NaNA), 2020, pp. 410-416, DOI: 10.1109/NaNA51271.2020.00076.

[11] Y. Zhang and Y. Feng, "Factors that Influence a Buyer's Decision Process of Shopping Online: The Effects of Tradition and Virtual Community," 2011 International Conference of Information Technology, Computer Engineering and Management Sciences, 2011, pp. 294-297, DOI: 10.1109/ICM.2011.316.

[12] X. Hu, Y. Yang, L. Chen and S. Zhu, "Research on a Prediction Model of Online Shopping Behavior Based on Deep Forest Algorithm," 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2020, pp. 137-141, DOI: 10.1109/ICAIBD49809.2020.9137436.

[13] H. Cheng and T. Fu, "The Determinants of Online Shopping Behavior," 2018 International Conference on Intelligent Autonomous Systems (ICoIAS), 2018, pp. 97-100, DOI: 10.1109/ICoIAS.2018.8494098.

APPENDIX A

A.1 Feasibility Assessment

The system has been highly feasible due to the fact that this is a graduate level approach that is being developed with the use of open source software such as libraries, integrated development environments such as NetBeans, Spyder etc. and other resources. This leads to a highly feasible approach defined in this research work.

A.1.1 Technical Feasibility

The approach defines in this research work is technically feasible as it does not contain a high, industry level coding. The opensource software such as OpenCv has been useful in interfacing the video input with the code easily. This is very useful in developing the approach technically without any kind of technical roadblocks due to the nature of the deployed approach.

A.1.2 Economical Feasibility

The prescribed approach is extremely cheap and economically feasible due to the lack of any extra resources such as high end cameras or expensive sensors that are required to achieve the prescribed approach. This is also helpful as students will not be having a lot of resources which makes this is suitable system for implementation.

A.1.3 Deployment Feasibility

The proposed methodology can be deployed easily as the deployment of the approach requires readily available resources to a student, such as the laptop and free or open source software such as Spyder IDE, Anaconda, Python SDK, Java SDK etc. and deployment resources.

A.1.4 Time Feasibility

The presented technique has been achieved in the required timeline as there was more than enough time provided to achieve the goals of the approach. The timeline for the development of the project has been reasonable and the workload was divided amongst the 4 members of the group which was useful for obtaining our approach easily.

A.2 NP Hard

A.2.1 Computations, Decisions and Languages

The most common resource to analyze software is time and number of execution steps, this is generally computed in terms of n. We will use an informal model of a computer and an algorithm. All the definitions can be made precise by using a model of a computer such as a Turing machine.

While we are interested in the difficulty of a computation, we will focus our hardness results on the difficulty of yes -no questions. These results immediately generalize to questions about general computations. It is also possible to state definitions in terms of languages,where a language is defined as a set of strings: the language associated with a question is the set of all strings representing questions for which the answer is Yes.

A.2.2 The Class P

The collection of all problems (Algorithms or methods that we are using in our project) that can be solved in polynomial time is called P. That is, a decision question is in P if there exists an exponent k and an algorithm for the question that runs in time $O(n^k)$ where n is the length of the input.

P roughly captures the class of practically solvable problems. Or at least that is the conventional wisdom. Something that runs in time 2^n requires double the time if one adds one character to the input. Something that runs in polynomial time does not suffer from this problem.

A.2.3 The Class NP

The collection of all problems that can be solved in polynomial time using non determinism is called NP. That is, a decision question is in NP if there exists an exponent k and an non deterministic algorithm for the question that for all hints runs in time O(n^k) where n is the length of the input.

A.2.4 P versus NP

It would seem that P and NP might be different sets. In fact, probably the most important Unsolved problems in Mathematics and Computer Science today is:

Conjecture. P $\neq$ NP

If the conjecture is true, then many problems for which we would like efficient algorithms do not have them. This would be sad. If the conjecture is false, then much of cryptographic under threat. Which would be sad.

A.2.5 NP Complete

While we cannot determine whether P = NP or not, we can, however, identify problems that are the hardest in NP. These are called the NP-complete problems. They have the property that if there is a polynomial-time algorithm for any one of them then there is a polynomial-time algorithm for every problem in NP.

A.2.5.1 Definition

A decision problem S is defined to be NP-complete if

a) It is in NP; and

b) For all A in NP it holds that A$\leq$PS.

A.2.5.2 Note that this means that

- If S in NP-complete and S in P, then P=NP.
- If S is NP-complete and T in NP and S≤PT, then T is NP-complete.

A.2.5.3 Example

We can state this even in simpler form, like as shown below:

Let us consider a module C_{NN} (Convolution Neural Network) in our system called S,

Then If C_{NN} is set to change in time T

$C_{NN}{}'$(changed module) will be the changed module

If ($C_{NN}{}' \in$ S)$\leq$ T

Then system is considered as NP Complete. Our System unconditionally satisfies this problem, So we can conclude our system as NP Complete.