HOME APPLIANCE CONTROL FOR VISUALLY AND VERBALLY IMPAIRED USING DEEP LEARNING AND IoT FOR GESTURE RECOGNITION

A PROJECT REPORT

Submitted by

MEENAKSHI. S. S	**0019121012**
RAKESH. K	**0019121016**

Under the guidance of

DR. P. SANKAR

in the partial fulfilment of the grant

for

THE STUDENT SCHEME

TAMIL NADU STATE COUNCIL FOR SCIENCE AND TECHNOLOGY

CHENNAI- 603 103

MAY 2023

BONAFIDE CERTIFICATE

Certified that this project report **HOME APPLIANCE CONTROL FOR VISUALLY AND VERBALLY IMPAIRED USING DEEP LEARNING AND IoT FOR GESTURE RECOGNITION** is the bonafide work of "**MEENAKSHI. S.S (19121012) and RAKESH. K (19121016)"** who carried out the project work under my supervision during the academic year 2022-2023.

SIGNATURE	**SIGNATURE**
Dr. P. Sankar	**Dr. A.L.Vallikannu**
Professor	Head of the department

ACKNOWLEDGEMENT

It is an extreme pleasure to thank our Chancellor **Dr Anand Jacob Verghese**, Pro Chancellor **Mr. Ashok Verghese**, Vice Chancellor **Dr. S.N. Sridhara**, Pro Vice-Chancellor **Dr. R.W. Alexander Jesudasan,** Registrar **Dr. Pon Ramalingam** of Hindustan Institute of Technology and Science for providing a conducive environment that helped us to pursue our project diligently and efficiently.

We wish to express our sincere gratitude to the Dean (Research and Development) **Dr. J. M. Mathana**, Hindustan Institute of Technology and Science for her valuable time, directions, suggestions, and support.

We are indebted to our Head of the Department **Dr. A.L. Vallikannu**, Electronics and Communication Engineering for her valuable guidance and technical support in the accomplishment of our project.

We would like to express our gratitude to our Project Guide **Dr. P. Sankar**, Professor of Electronics and Communication Engineering for his relentless guidance and efforts in completing our project.

Our sincere thanks to all the teaching, non-teaching staff and family members who have been constantly supporting us throughout the accomplishment of this project.

HINDUSTAN
INSTITUTE OF TECHNOLOGY & SCIENCE
(DEEMED TO BE UNIVERSITY)
TNSCST

ABSTRACT

Emerging technologies make human life easier by giving us the greatest comfort at our fingertips. Yet the accessibility of these smart devices by patients with visual or verbal impairments is time-consuming. For instance, Smart home automation devices such as Alexa or Siri require vocal activation, which is impossible for a mute person. This proposed device uses Gesture recognition as an alternative to vocal activation. Gesture recognition techniques have been used for small-scale applications of home automation, but a model exclusive for impaired patients has always only been theoretical. Embedding this deep learning-based gesture recognition with an IoT module helps us save the data in the cloud and aid with easy simultaneous accessibility for the impaired. Hence, This proposed module is a small watch-like device that employs hand gesture recognition using CNN techniques for appliance control and uploads the obtained data to the server which can be utilized by any impaired person to control any device connected to the local network without any external aid. This module can be tested in a homestay for the impaired to make the stay more technologically viable.

TABLE OF CONTENTS

TABLE OF FIGURES

HINDUSTAN INSTITUTE OF TECHNOLOGY & SCIENCE (DEEMED TO BE UNIVERSITY)

TNSCST

TAMIL NADU STATE COUNCIL FOR SCIENCE AND TECHNOLOGY

DOTE CAMPUS, CHENNAI- 600 025

THE STUDENT SCHEME

UTILIZATION CERTIFICATE

1. **Name of the Guide and Address**: Dr. Sankar. P
   (Professor)
2. **Name of the Student (s):** Meenakshi. S.S
   Rakesh. K
3. **Title of the Project**: Home Appliance Control for Visually and Verbally Impaired using Deep Learning and IoT for Gesture Recognition.

It is certified that a sum of Rs ( ) sanctioned by the council for carrying out the above-mentioned student project has been utilized for the purpose for which it has been sanctioned and the sum of Rs unutilized is refunded.

SIGNATURE	**SIGNATURE**	**SIGNATURE**	**SIGNATURE**
Dr. P. Sankar	**Dr. A.L.Vallikannu**	**Dr.J.M. Mathana**	**Dr. Pon Ramalingam**
Professor	Head of the department	Dean (R&D)	Registrar

1 INTRODUCTION

The 21st century has witnessed the digitalization of all things around us. All appliances can be electronically managed in the comfort of one place using fore-fronting applications such as Deep learning and the Internet of Things (IoT). Controlling smart home appliances is a fore-fronting application of deep learning. Existing smart appliances are formulated locally adhering in mind thetwo assumed facts:

- Any kind of controlling of the appliance can be attained by vocal activation which is a hindrance for a verbally impaired person.
- The user should stay in the vicinity of the device they wish to control which is a major con for a visually impaired person.

An alternative to control smart home appliances rather than by speech recognition is gesture recognition. In this report, a small wearable watch-like device that a verbally or visually impaired person can use to control an appliance using hand gestures is proposed. The primary focus of the project is to provide easy accessibility of smart devices to visually and verbally impaired people by making the interface between the user and the device wireless. The secondary focus of this project is to make a cost-efficient, compact and portable device that any impaired person can use to control any smart device at any point in time and place.

The prerequisite behind any smart device is reading and understanding the gesture and corresponding it to its allotted functions. The device aims to achieve this by using an IoT camera ESP-32 which is connected to a common network as the appliance the user wants to control. The camera reads the associated gestures of the person using a deep learning

Convolutional Neural Networks (CNN), enables control of the appliances according to the given gestures using the client-server relationship for the Internet of Things (IoT). As a proof of concept, the project is implemented for just one light source whose results are consolidated in the chapter.

However, a few issues that we hope to resolve include Power consumption since the projecthelps assist impaired people, it requires having to stay ON state for long hours. This may cause additional power usage.

1.1 AIM:

To design and build a wearable IoT-based hand gesture recognition device that can recognize hand gestures using Convolutional Neural network and TensorFlow Lite to aid verbally or visually challenged users in home appliance control.

1.2 OVERVIEW OF THE REPORT:

The report is organized as follows: The previously emerged works regarding the methods for gesture recognition is being mentioned briefly in Chapter 2 as Background Research and the research gap in previous works that gave light to this proposed device is also elaborated. Chapter 3 sheds light on the proposed device, the hardware and the software components required and the working mechanisms of the algorithms used. The device Flowchart and the layer of the CNN illustrations are explained in this chapter as well. Furthermore, the Result and Analysis of the module are shown in Chapter 4 with real-time images of the output along with a Comparative graph to show the Performance evaluation of the proposed device for gesture recognition in comparison to the other methods. The final chapter deals with the Conclusion and other discussion of the device. The report ends with the reference index of all the papers that we have referred to.

2 BACKGROUND RESEARCH

Deep Learning applications for gesture recognition have been the talk of the last few years. An extensive background analysis of these methods along with their methodologies, limitations and overall research gap is discussed in this chapter.

2.1 LITERATURE SURVEY:

The existing systems help us in using vocal recognition techniques for appliance control which is a tedious process to be used by a mute person [12]. Hence wearable devices within proximity of the appliance were developed. This was a hindrance for blind people as they could not be within the proximity of the appliance they wish to operate [10].

In a research paper by Ali Moin and Andy Zhou, the potential application of wearable devices that utilize surface electromyography to monitor muscle activity in the development of hand gesture recognition applications is discussed. These devices commonly employ machine learning models, either locally or externally, for the classification of gestures. However, it is noted that most devices with local processing cannot train and update the machine-learning model while in use, leading to a less-than-optimal performance in real-world conditions [1].

Finally, deep learning methods have been coalesced with other FGR models for the implementation of real-time and dynamic hand gesture recognition systems which use different techniques to extract hand-crafted features followed by a sequence modelling technique. Firstly, Antenna systems were embedded along with deep learning algorithms to provide point-to-point communication between the user and the devices in a paper proposed by Sruthy et al where they used a miniature radar sensor to capture Doppler signatures of 14 different hand gestures and train a deep convolutional neural network (DCNN) to classify these captured gestures. The classification results of the proposed architecture show a gesture classification accuracy exceeding 95% and a very low confusion between different which is almost a 10% improvement over the single-channel Doppler methods reported in the literature gestures however, it declined as the distance between the gesture and the camera is reduced(zoom-in) [2].

Since EMG showed optimism for FGR, The sensing and Classification based FGR model was proposed by Jiantin et al where the Sensing was carried out by EMG signals and the Classification was carried out by a well-trained CNN algorithm. The thesislacked to show its adoption with latent factor analysis, cognitive computing, and attention mechanism [3].

Later, EMG-based FGR was imbibed to deep learning which was a model proposed by Ali et al. It was observed that regardless of network configuration some motions (close hand, flex hand, extend the hand and fine grip) performed better (83.7% ± 13.5%, 71.2% ± 20.2%, 82.6% ± 13.9% and 74.6% ± 15%, respectively) throughout the course of study. So, a robust and stable myoelectric control can be designed based on the best-performing hand motions. With improved recognition and uniform gain in performance, the deep learning-based approach has the potential to be a more robust alternative to traditional machine learning algorithms [5].

Since it is established that deep learning algorithms along with other methods of finger recognition yield the best accuracy rates, different EMG- based recognition techniques using various techniques such as ANN, SVM, RF and LR and their statistical comparisons were performed by Kyung et al where classification using ANN showed the most accuracy by just considering the TD features. ANN was confirmed to be the least affected by individual variability. Future works include a more heterogeneous population to see the extent to which the ANN algorithm holds good for recognition [6].

Hence, Zhi-Hua et al contributed to work with finger segmentation algorithms by sending real-time gestures captured using a camera to CV frameworks. Its accuracy depended on its ability to ignore background frameworks and detect the palm and fingers separately. This module addressed all the cons in the above-mentioned papers but lacked to show a reliable accuracy rate when implemented [7].

The innovation of computer vision techniques was briefly explored by Munir et al. CV was introduced for Hand gesture recognition as they did not utilize robust devices clung to the body, making this technique the most optimistic one for elderly people as wearables cause discomfort to age-old users [8].

Later, Deep Learning based gesture recognition models were introduced as DL showed optimistic results for image recognition. In the thesis proposed by Jinxian et al the principal component analysis method and GRNN neural network are used to construct the gesture recognition system which reduced the redundant information of EMG signals, reduced the signal dimension, improved the recognition efficiency and accuracy, and enhanced the feasibility of real-time recognition. After dimensionreduction and neural network learning, the overall recognition rate of the system reached 95.1%, and the average recognition time was 0.19s [9].

In their study, Andrés G. Jaramillo and Marco E. Benalcázar utilize surface electromyography (EMG) and Machine Learning techniques. The aim is to identify gestures through EMG signals, which is a challenging task due to the various physiological processes involved in the generation of skeletal muscle movements. Limitations arise in terms of the number of gestures that can be recognized (referred to as classes) and the time required for processing. Hand gesture recognition models that employ EMG and Machine Learning face two primary difficulties: noise in the EMG signal and the relatively small number of gestures per person compared to the amount of data generated by each gesture, known as the curse of dimensionality [10].

CV methods showed high compatibility when compared to EMG **Doppler-based** FGR methods as reviewed in paper 7 hence Chen et al proposed to use CV with CNN to recognize gestures and to have more accessibility an IoT module was introduced to employ control from anywhere. Though it succeeded to access appliance control from anywhere, users were still required to carry huge and wired parts around them. Hence, this paper shed light on our proposal [12].

2.2 RESEARCH GAP:

In this project, we hope to fulfil the void in the research gap by addressing the above-mentioned statements by utilizing the following:

- A small wearable watch like an IoT-based device for gesture recognition using Deep Learning methods with no robust sensors.
- Uses the TF LITE model as a framework to make the recognition time spontaneous.

3. SYSTEM PROPOSAL

A device worn by challenged users is proposed to overcome the mentioned research gaps from chapter two. The internal stages behind the watch-like hardware are explained in terms of its workflow.

3.1: STAGE 1: The general workflow is illustrated in Figure 3.1 starting with an ESP-32 camera which is used to capture real-time images of any gesture made and actively transmits the captured image to a server. A vibrating motor is used to indicate to the impaired person if their gesture is read and recognized or not. If it generates an output in the form of vibration, the gesture is recognized if not, the camera captures the gesture again. Furthermore, the captured image is flipped to ensure consistency, augment training data, handle ambidextrous gestures, and facilitate mirror reflections when necessary.

3.2: STAGE 2: The flipped image is now read to see if a palm is found using the Hand Recognition Algorithm. First, the neural network searches for the presence of the user's hand in the image feed. Later, the background is eliminated to crop an image covering just the palm.

3.3: STAGE 3: Once a cropped palm is obtained, another neural network, the 'Key Point Detection algorithm' is activated which aids in identifying key points of the palm and returns the best features for the recognition of gestures. Furthermore, the points are solved using a PnP matrix which converts all the 3-D vector images into a 2-D vector plane.

3.4: STAGE 4: These key points help us in determining a reference point in the palm such that all the vector distances from the reference point are measured using the Euclidean distance formula:

EUCLIDEAN DISTANCE (d) = sqrt [(x2 – x1) ^2 – (y2 – y1)^2]

to identify the user's finger signs using the 'Finger gesture Recognition algorithm'

3.5: STAGE 5: Any appliance connected to the local network of the internet camera module is eligible to send an API request for pairing the devices with the server. Based on previously paired numbers to appliance data, that specific appliance is activated using the IoT platform. After authorizing the request received, the appliances generate an output function concerning the gesture pointed at the camera module. This way any impaired person can operate and control any device connected to the local network by associating the gesture with the function they want to perform. Figure 3.2 illustrates the device worn by the user on the wrist. It consists of an ESP-32 camera along with the TTL programmer which are both strapped in the elastic watch-like band.

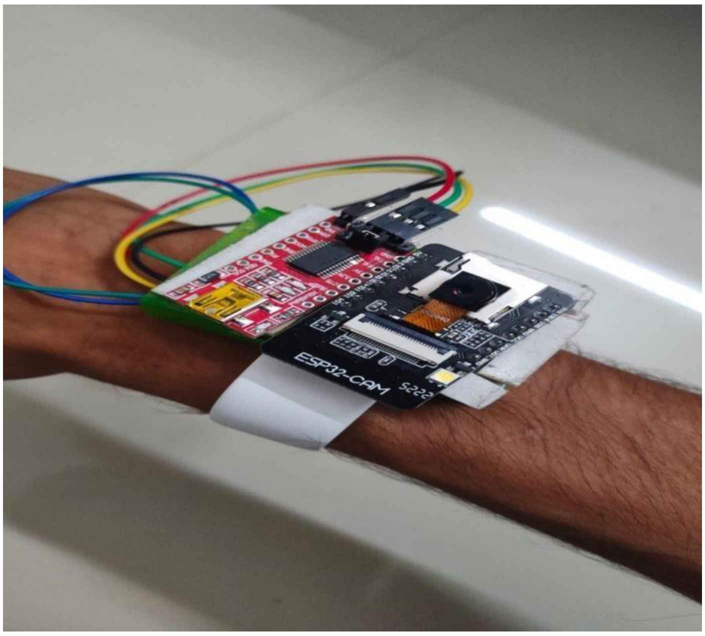

Figure 3.2: the top perspective of the device

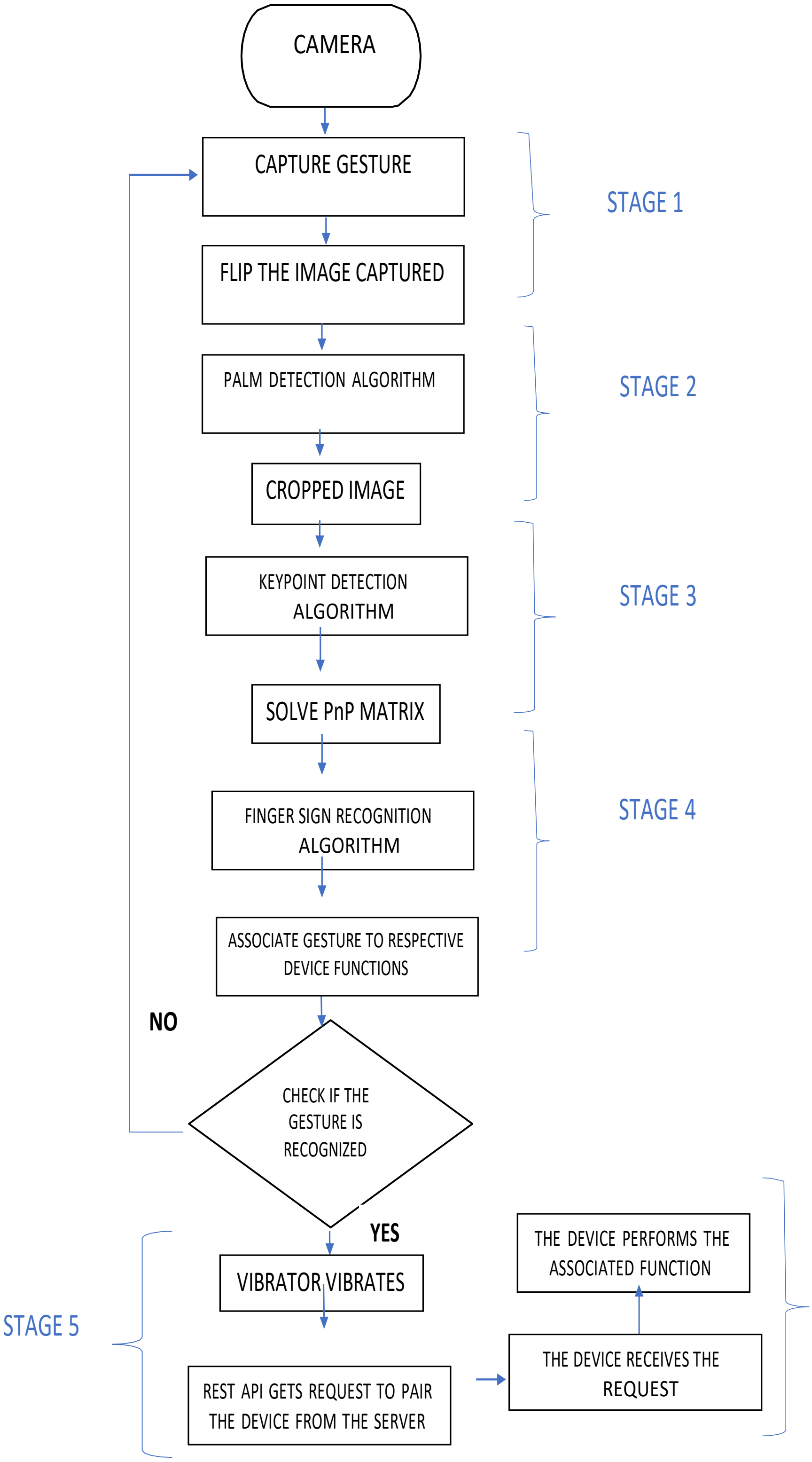

Figure 3.1: flowchart of the proposed model

4. RESULTS AND ANALYSIS

The algorithms explained have been trained and tested for real-time detection and recognition of gestures. The output obtained from undergoing the stages from Figure 3.2 is illustrated in Figure 4.1 whose output is later used for appliance control by giving invoking the client-server relationship of the Rest API.

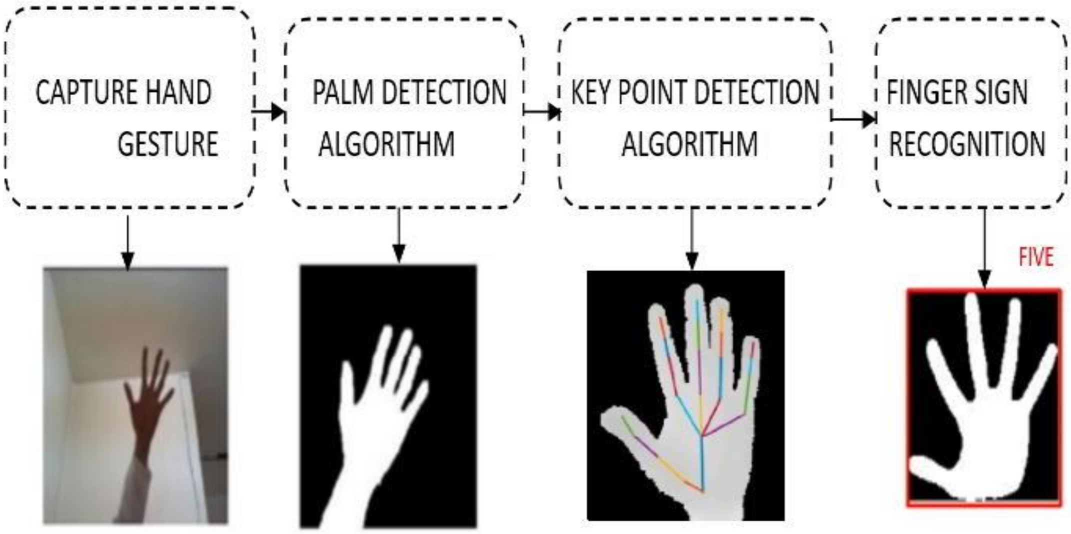

Figure 4.1: Specific Algorithmic Output

The proposed model has been simulated to yield appropriate results in real-time applications. As a proof-of-concept evaluation of the source code is run to use gestures to Switch a light source on and off and to control the audio input of another device. Figure 4.2 illustrates the gesture number "TWO" being read, recognized and labelled.

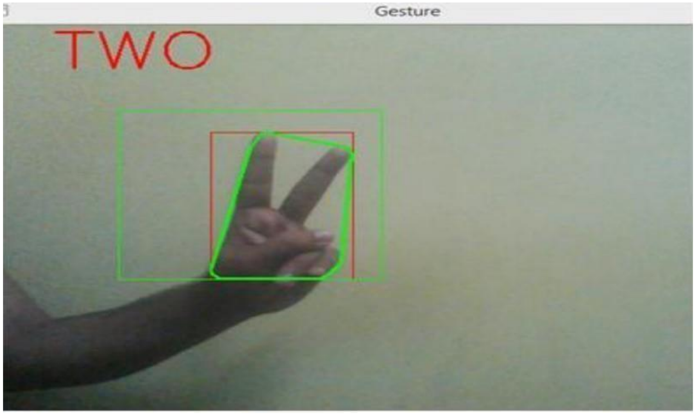

Figure 4.2: Gesture number "2" shown by the user

Figures 4.3 and 4.4 below show the verified output of a system being able to recognize a gesture posed and associate it with the controlling of a LED (a feasible light source). As a proof-of-concept evaluation of the source code is performed to use gestures to Switch a light source on and off and in Figure 4.5 to control the color of the LED.

Figure 4.3: The User gestures "1" to switch ON the LED

Figure 4.4: User gestures "2" to switch OFF the LED

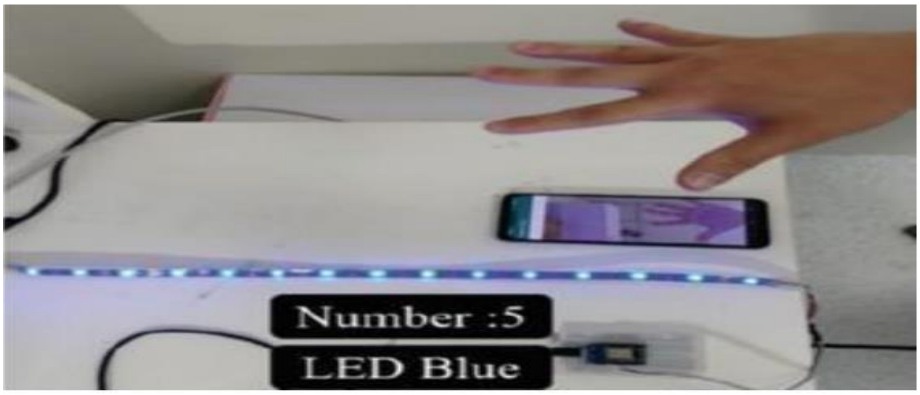

Figure 4.5: User gestures "5" to change LED colour from RED to BLUE

To analyze the accuracy of the proposed algorithm, training was carried out for 6 gestures made by impaired people. A self-made dataset with a total of 240 images was collected and used for training/testing purposes. The dataset consists of 6 classes:- “1”,”0”,”3”,”5”,”6” and ”7” which were the basic few gestures set up for pairing up with appliances. Each class consisted of 60 images out of which, 40 images were used for training purposes and 20 images were used for testing purposes. As a whole, 240 images were used for training the neural network and 120 images were used for testing purposes. The confusion matrix(CM) is derived in Table 1 highlighting the diagonals which represent the true positives. From the CM the True positive(TP), True negative(TN), False positive(FP) and False negative(FN) values is comprehended as follows:

TP final = 92, TN final = 458, FP final = 28 , FN final= 20

The accuracy, recall and precision values are calculated from the above-obtained parameters from the formula

$$\textit{Recall} = \frac{TP}{TP + FN} \quad \textit{Precision} = \frac{TP}{TP + FP} \quad Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The recall value rounded to 82% while the precision value rounded to 78%. Our algorithm showed decent accuracy which is approximately 91.97%. The model's recall and precision values can be boosted by adding more intra-class datasets.

Predicted label

Actual

	Class '1'	Class '0'	Class '3'	Class '5'	Class '6'	Class '7'
Class '1'	13	3	0	1	0	3
Class '0'	0	16	1	1	0	2
Class '3'	0	1	14	0	0	0
Class '5'	0	0	0	18	2	0
Class '6'	1	0	0	0	15	4
Class '7'	6	0	0	0	3	11

Table 1: confusion matrix of the proposed algorithm

Different gesture recognition techniques were also tested during the process Figure 4.6 illustrates how CNN worked best with the device showing an accuracy of 91% as it is known for its ability to capture spatial features and patterns which makes them highly effective in recognizing complex gestures accurately. Followed by RNN (Recurrent Neural Network) with 87% accuracy as they are suitable for sequential data analysis which makes them useful for recognizing gestures that involve temporal patterns or dynamics. When gesture recognition was implemented with SVM (support vector matrix) 79% accuracy was obtained as SVMs struggled with more complex and high-dimensional gesture recognition tasks. HMM (Hidden Markov Model) is well-suited for modelling sequential data which makes them appropriate for gesture recognition with an accuracy rate of 75%. RF(Random Forest) is the only ensemble learning method that is tested with our device which generated 81% accuracy. While it performs well in various domains, its performance is lower compared to specialized models like CNNs for gesture recognition.

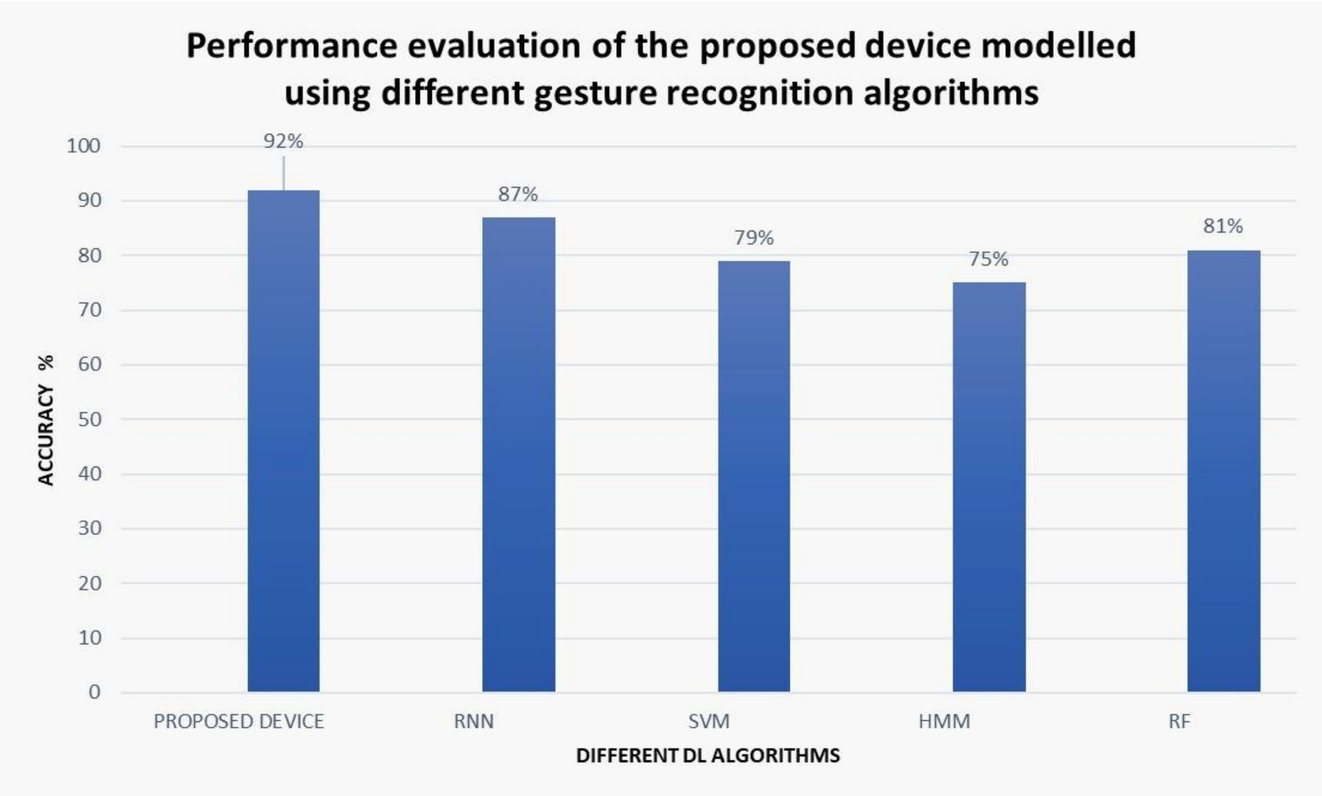

Figure 4.6: Comparative Analysis Chart

5. CONCLUSION AND DISCUSSIONS

The final chapter of the report consolidates all the findings of the executed module and puts forth the conclusion and discussions regarding future advancements based on its final findings. For a few demerits mentioned in the previous chapter of the report, theoretical solutions have been provided in the advancements part.

5.1 SUMMARY OF THE PRESENTED WORK

In this work, a model capable of recognizing and associating gestures with respective controlling functions of small-time appliances is consolidated. previously published papers and existing projects that utilize gesture recognition using various technologies, along with their drawbacks have been briefly discussed. The skeletal outline of the proposed model shows the data capturing, data manipulation and output representation. Furthermore, details about the intricate workings of the proposed model are elucidated Gesture recognition is done so with the help of a convolutional neural network model which was trained using Python and embedded with TTL PROGRAMMER and Tensor Flow Lite model. Output and analysis were tested using an LED light source for ON, OFF and COLOR CHANGE functions.

5.2 DISCUSSIONS

The device has shown to be more effective and efficient than all the other models described in the literature survey from Chapter 2. Some of the advantages observed that give this device front over the others is described below:

- This proposed device helps in unmonitored control of appliances for impaired people sinceit only uses simple gestures as a merit of control
- It has shown to be highly cost-effective as it uses the concept of IoT and needs only proper network connectivity for efficient managing of data. Due to the usage

of local connectivity, the model implements immediate and powerful interaction with high Speed and sufficient reliability for the recognition system.

- For capturing and recognizing real-time images, the model shows a good performance system with complex background and showcases fast and powerful results from the proposed algorithm

On the other hand, just like every other technologically accessible model, the implemented model along with its algorithm shows its weaknesses. Most limitations include weak discoverability of the devices in the local network due to poor signal strength. Furthermore, an Irrelevant object might overlap with the hand of the user that raises a gesture. The performance recognition algorithm decreases when the distance greater than 1.5 meters between the user and the camera is.

5.3 FUTURE ADVANCEMENTS

Using gesture recognition, good outcomes can be produced for various smart applications. Further scopes for this model include expanding this into a smart home control for the visually and verbally impaired with a cloud server specific just for the home network and making that model as cost-efficient as possible.

More applications of gesture recognition can be embedded with this such as smart wheelchairs and smart braille systems which operate concerning the association of a gesture made to its assorted function.

REFERENCES – 6

[1] Moin, A., Zhou, A., Rahimi, A. *et al.* A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition. *Nat Electron* **4**, 54–63 (2021).

[2] Jiajun Zhang, Jinkun Tao, Jiangtao Huangfu, Zhiguo Shi:
Doppler-Radar Based Hand Gesture Recognition System Using Convolutional Neural Networks. CoRR abs/1711.02254 (2017)

[3] Fu J, Cao S, Cai L and Yang L (2021) Finger Gesture Recognition Using Sensing and Classification of Surface Electromyography Signals with High-Precision Wireless Surface Electromyography Sensors. Front. Comput. Neurosci. 15:770692. doi: 10.3389/fncom.2021.770692

[4] Al-Hammadi, Muneer H. et al. “Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation.” *IEEE Access* 8 (2020): 192527-192542.

[5] Asif, Ali Raza, Asim Waris, Syed Omer Gilani, Mohsin Jamil, Hassan Ashraf, Muhammad Shafique and Imran Khan Niazi. “Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG.” *Sensors (Basel, Switzerland)* 20 (2020): n. pag.

[6] Lee, K.H.; Min, J.Y.; Byun, S. Electromyogram-Based Classification of Hand and Finger Gestures Using Artificial Neural Networks. Sensors 2022, 22, 225. https://doi.org/ 10.3390/s22010225

[7] Chen, Zhihua, Jung-Tae Kim, Jianning Liang, Jing Zhang and Yubo Yuan. "Real-Time Hand Gesture Recognition Using Finger Segmentation." *The Scientific World Journal* 2014 (2014): n. pag.

[8] Oudah, Munir, Ali Al-Naji and Javaan Singh Chahl. "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques." *Journal of Imaging* 6 (2020): n. pag.

[9] Qi, Jinxian, Guozhang Jiang, Gongfa Li, Ying Sun and Bo Tao. "Surface EMG hand gesture recognition system based on PCA and GRNN." *Neural Computing and Applications* 32 (2019): 6343-6351.

[10] Jaramillo-Yánez, Andrés, Marco E. Benalcázar and Elisa Mena-Maldonado. "Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review." *Sensors (Basel, Switzerland)* 20 (2020): n. pag.

[11] Tian, Yong, Chuanzhen Zhuang, Jiadong Cui, Runjie Qiao and Xuejun Ding. "Gesture recognition method based on misalignment mean absolute deviation and KL divergence." EURASIP Journal on Wireless Communications and Networking 2022 (2022): 1-21.

[12] Cheng-Ying Yang, Yi-Nan Lin, Sheng-Kuan Wang, Victor R.L. Shen, Yi-Chih Tung, Frank H.C. Shen & Chun-Hsiang Huang (2023) Smart Control of Home Appliances Using Hand Gesture Recognition in an IoT-Enabled System, Applied Artificial Intelligence, 37:1, 2176607, DOI: 10.1080/08839514.2023.2176607

[13] Skaria, Sruthy, Akram Al-Hourani, Margaret Lech and Robin J. Evans. “Hand-Gesture Recognition Using Two-Antenna Doppler Radar With Deep Convolutional Neural Networks.” *IEEE Sensors Journal* 19 (2019): 3041-3048.

[14] Ciaburro G, Padmanabhan S, Maleh Y, Puyana-Romero V. Fan Fault Diagnosis Using Acoustic Emission and Deep Learning Methods. *Informatics*. 2023; 10(1):24. https://doi.org/10.3390/informatics10010024