GESTURE AND VOICE RECOGNITION SYSTEM USING DEEP LEARNING TO CONTROL IOT APPLICATIONS: A LITERATURE SURVEY

[1]R Mohit Sai Reddy, [2]Sharath B, [3]Dr. Jagruthi H

[3]Associate Professor
Information Science and Engineering
B.N.M. Institute of Technology, Bengaluru

Abstract- **This literature survey delves into the realm of gesture and voice recognition technologies tailored for controlling Internet of Things (IoT) devices in smart home environments. With the proliferation of IoT devices, the demand for intuitive and efficient control mechanisms has surged, leading to the exploration of gesture and voice recognition as viable solutions. The survey examines the landscape of existing research, focusing on the utilization of deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), for enhancing gesture and voice recognition accuracy. It highlights the evolution of these technologies, their applications in smart home environments, and the challenges faced, including environmental robustness and user privacy. By critically analyzing the strengths and limitations of current methodologies, the survey aims to provide insights into the future of gesture and voice recognition for IoT device control, identifying potential areas for further research and development.**

1. INTRODUCTION

This literature survey investigates the field of gesture and voice recognition for controlling Internet of Things (IoT) applications, focusing on the integration of deep learning techniques. Gesture and voice recognition offer intuitive interfaces for interacting with IoT devices, but challenges exist in achieving high accuracy and responsiveness. Deep learning, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), presents an opportunity to improve recognition accuracy and system responsiveness.

The project aims to develop a comprehensive system using deep learning for gesture and voice recognition, enhancing the control and interaction with IoT devices. By leveraging CNNs for gesture recognition and RNNs for voice recognition, the system seeks to achieve higher accuracy and efficiency in recognizing user inputs. Integration with IoT devices enables seamless control through gestures and voice commands, offering a more intuitive user experience.

Through this literature survey, the goal is to explore existing research and methodologies in gesture and voice recognition, particularly focusing on deep learning approaches. By analyzing the strengths, limitations, and comparative efficacy of these techniques, the survey aims to provide insights into improving the accuracy and responsiveness of gesture and voice recognition systems for controlling IoT applications. Additionally, it aims to identify gaps and propose potential avenues for future research, aiming to advance the field of gesture and voice recognition in IoT applications.

2. LITERATURE SURVEY

In [1], the implementation of hand gesture recognition (HGR) systems involves several key components, including segmentation, feature extraction, and classification stages. In the proposed system by Parvathy et al., a vision-based approach using machine learning techniques was employed to achieve high accuracy in gesture recognition. The system utilized techniques such as Discrete Wavelet Transform (DWT) for extracting rotation and scale invariant key descriptors, followed by the Bag of Word technique to develop fixed dimension input vectors for the Support Vector Machine (SVM) classifier. The classification accuracy of specific gestures, such as 'No' and 'grasp', reached an impressive 98%, with an overall classification accuracy of 96.5% and a recognition time of 0.024 seconds. Despite the success of the proposed HGR system, several challenges exist in the development and implementation of such systems. One of the primary challenges is the complexity of background environments, which can affect the accuracy of gesture recognition. Ensuring robustness to varying backgrounds and lighting conditions is crucial for the system to perform effectively in real-world scenarios. Another challenge lies in efficient feature extraction and classification algorithms. The effectiveness of an HGR system heavily relies on the ability to extract relevant features from input data and classify

gestures accurately. Developing algorithms that can handle the variability in hand gestures, such as different hand shapes and movements, poses a significant challenge in achieving high recognition accuracy. Furthermore, the integration of real-time processing capabilities in HGR systems presents technical challenges in terms of computational efficiency and speed . Ensuring that the system can process and recognize gestures in real-time without compromising accuracy is essential for providing seamless interaction experiences for users.

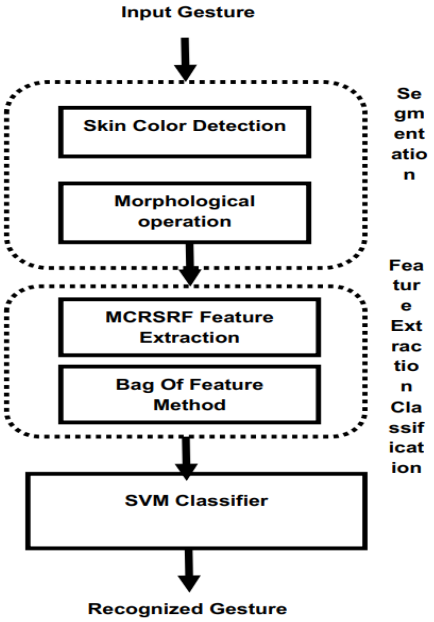

Figure 1. Flowchart of proposed HGR System

In [2], the author introduces a methodology that revolutionizes hand gesture recognition through the utilization of convolutional neural networks (CNN). The implementation of this methodology involves a multi-step process starting with the segmentation of the hand region of interest from input images using advanced image processing techniques. Subsequently, features are extracted from the segmented hand region to capture essential information for gesture classification. These features are then inputted into a CNN classification model that has been trained on a dataset of hand gesture images. The CNN model leverages its hierarchical layers to learn discriminative features and accurately classify the hand gestures, showcasing the power of deep learning in enhancing human-computer interaction systems. Despite the promising results achieved by this methodology, several challenges were encountered during its development and implementation. One significant challenge lies in the variability of hand gestures across different individuals, poses, and environmental conditions, which can lead to difficulties in accurately detecting and recognizing gestures in real-world scenarios. Additionally, the methodology requires a large and diverse dataset for training the CNN model, posing a challenge in terms of data collection and annotation efforts. Furthermore, optimizing the CNN architecture and hyperparameters to achieve the desired performance levels demands expertise and computational resources. These challenges underscore the complexities involved in developing robust hand gesture recognition systems and highlight the importance of addressing these issues to further refine the methodology's effectiveness and applicability in practical settings.

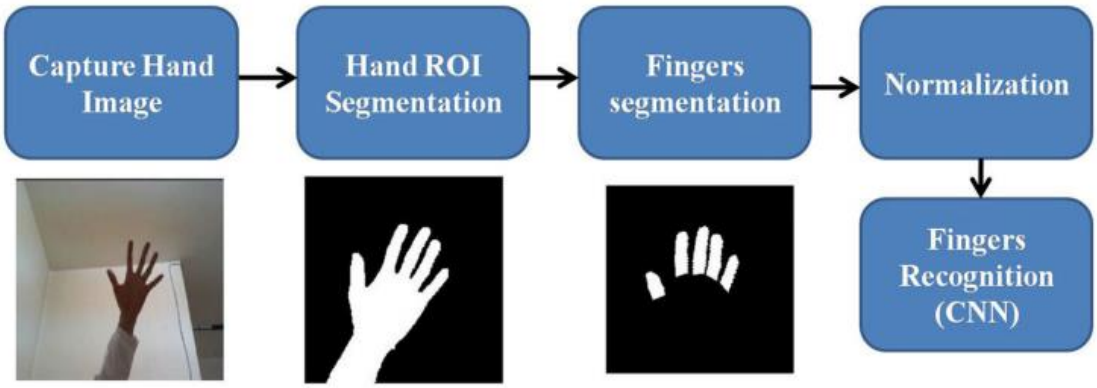

Figure 2: Proposed Hand Gesture Recognition System

In the realm of smart home technology, the integration of natural language processing (NLP) and voice recognition with Internet of Things (IoT) technologies has emerged as a promising avenue for enhancing user interaction and control.

The implementation presented by the authors in [3] showcases a novel approach to smart home management. This implementation, known as the IoT Agent, leverages third-party APIs such as the Dialogflow API for NLP, the Web Speech API for voice recognition, and the Firebase API for storage, highlighting a multi-tier architecture that streamlines the development of IoT applications. One of the key challenges addressed in this implementation is the seamless integration of diverse third-party services to create a cohesive and efficient system. By utilizing APIs from different providers, the IoT Agent demonstrates how modern IoT applications can leverage existing technologies to deliver robust functionality. This approach not only simplifies the development process but also ensures that the system remains up to date with the latest advancements in NLP, voice recognition, and cloud storage solutions. However, the integration of multiple APIs also introduces complexities in terms of system compatibility, data synchronization, and maintenance, requiring careful attention to ensure smooth operation.

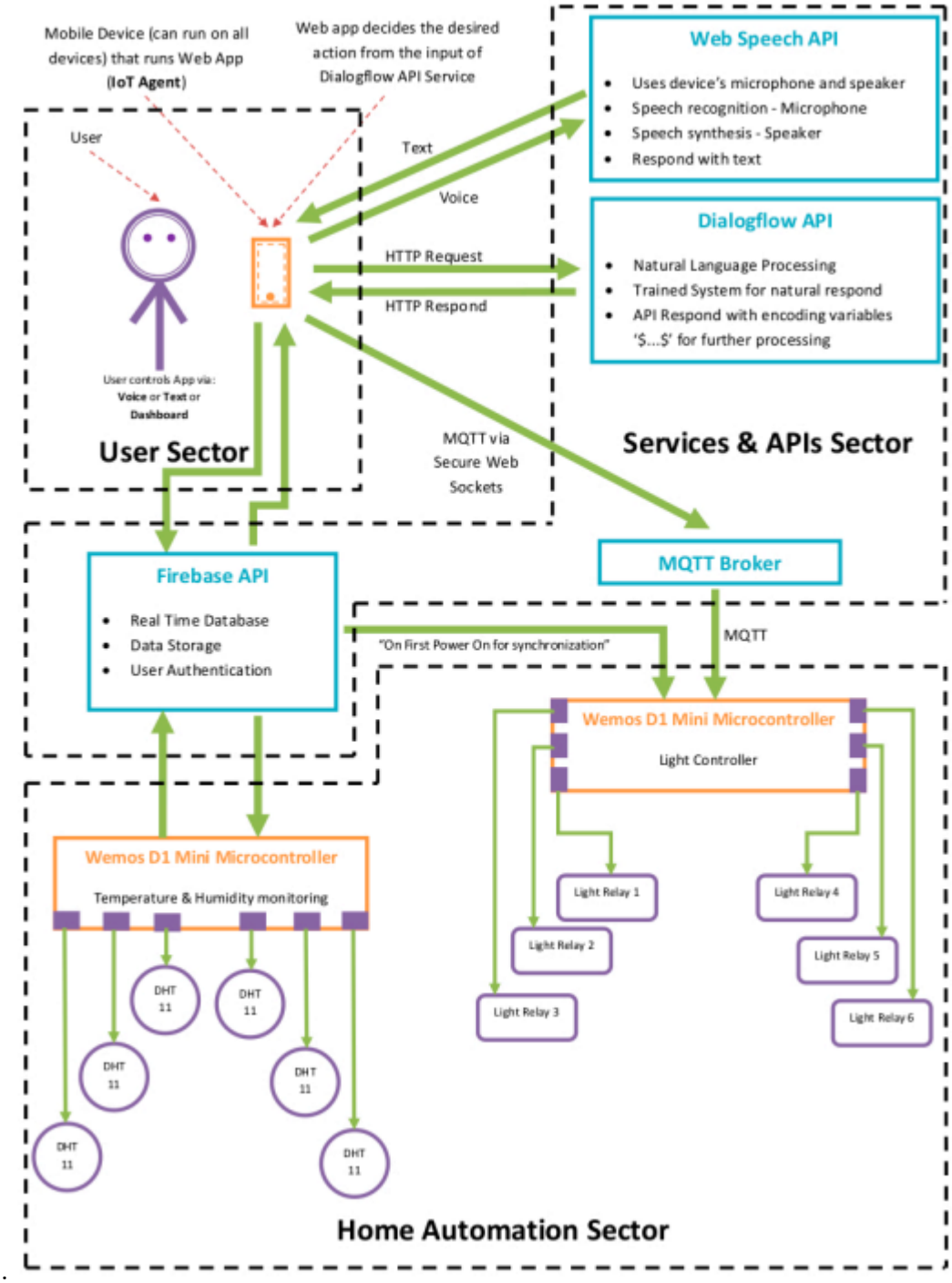

Figure 3. Main System Architecture

In [4], the voice controlled smart home automation system proposed by the authors integrates artificial intelligence (AI) and internet of things (IoT) technologies to enhance the control and efficiency of smart home systems. This system allows users to interact with their home appliances and devices through voice commands, providing a convenient and

intuitive way to manage various functions within the household. By leveraging AI and IoT, the system aims to streamline the user experience and improve overall home automation capabilities. In terms of implementation, the system is divided into five key sections. The first section introduces the study and sets the context for the integration of IoT and AI in smart home automation. The second section reviews relevant works in the field, highlighting the importance of voice recognition technology and natural language processing (NLP) in enhancing user interactions with smart home systems. The third section details the components of the proposed system, outlining how IoT devices and AI algorithms work together to enable voice-controlled functionalities for home automation. One of the main challenges addressed in this paper is the seamless integration of voice recognition technology with IoT devices to ensure accurate and reliable control of home appliances. Implementing a robust NLP-based AI system that can accurately interpret and execute voice commands presents a significant technical challenge, requiring sophisticated algorithms and machine learning models. Additionally, ensuring the security and privacy of user data and voice interactions in a smart home environment is crucial, as voice commands may contain sensitive information that needs to be protected from unauthorized access. The implementation of the voice controlled smart home automation system involves deploying IoT devices throughout the home and connecting them to a central control system powered by AI algorithms. Users interact with the system by issuing voice commands, which are processed and executed by the AI engine to control various home functions such as lighting, temperature, and entertainment systems. The system's ability to understand and respond to natural language commands adds a layer of convenience and accessibility for users, making home automation more user-friendly and intuitive.

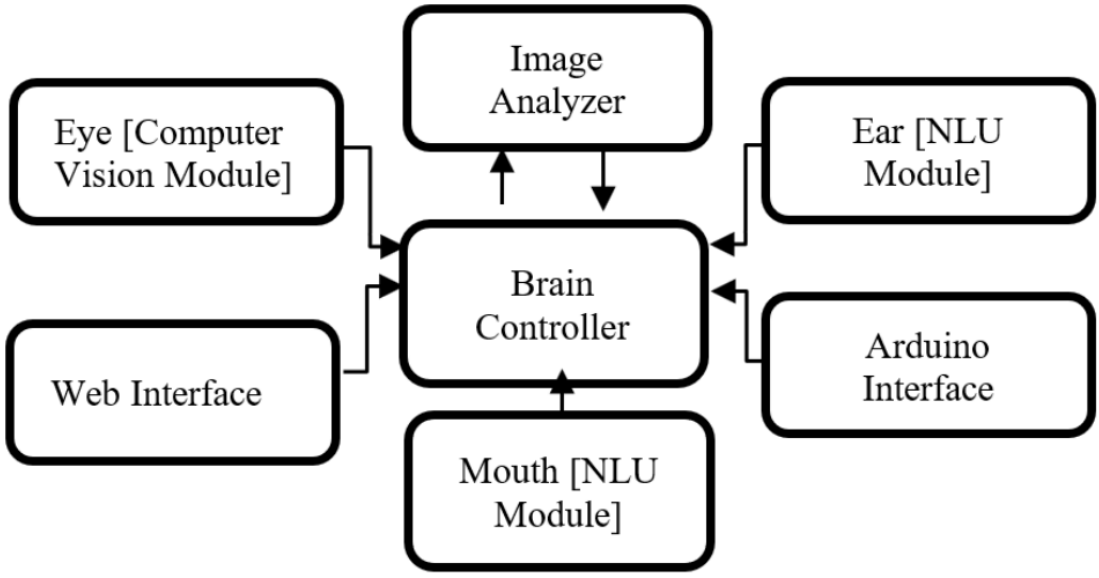

Figure 4. The proposed system architecture

The Web-Based Gesture Recognition System for Controlling Heterogeneous IoT devices using Deep Learning presents a novel approach to enhancing user interaction with IoT devices through gestures. By leveraging deep learning techniques, users can control multiple IoT devices using intuitive gestures, eliminating the need for individual mobile applications for each device. This system offers a user-friendly web interface that allows users to define and train gestures for specific IoT device actions, providing a seamless and efficient control mechanism. The implementation of this system involves two main phases: Training and Testing. During the Training phase, users train the gesture recognition classifier by performing various gestures associated with controlling IoT devices. These gestures are then used to initiate specific actions on the IoT devices. In the Testing phase, the web user interface recognizes the user's gestures and sends requests to the corresponding IoT devices, enabling seamless control through intuitive gestures. Despite the promising capabilities of this system, several challenges need to be addressed. One significant challenge is the need for sufficient training data to ensure accurate gesture recognition. Deep learning models typically require large amounts of data for training, which can be a bottleneck in real-world applications. Additionally, ensuring the robustness and reliability of gesture recognition across different users and environmental conditions poses a challenge that needs to be carefully addressed in the system design. Furthermore, the integration of diverse IoT devices with varying hardware setups into a common web interface introduces complexities in ensuring compatibility and seamless communication. The system must be designed to accommodate different types of IoT devices and their specific control requirements while maintaining a user-friendly and intuitive interface. Addressing these challenges will be crucial in realizing the full potential of the Web-Based Gesture Recognition System for controlling heterogeneous IoT devices using deep learning, paving the way for enhanced user experiences and efficient IoT device management.

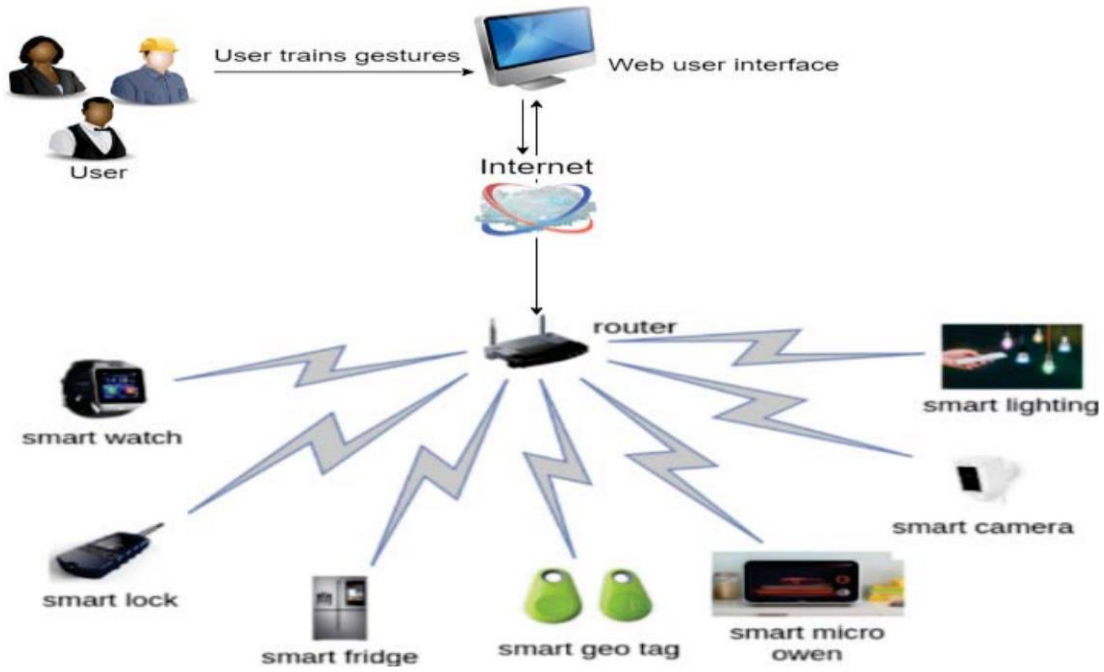

Figure 5. System Architecture

REFERENCES:

[1] D. Priyanka, P. Parvathy, K. Subramaniam, P. V. G. K. D., K. Kumar, J. Varghese, and J. Thangaiyan, "Development of hand gesture recognition system using machine learning," Journal of Ambient Intelligence and Humanized Computing, vol. 12, 2022.

[2] P. S. Dr-Neethu, S. Ramadass, and D. Sathish, "An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks," Soft Computing, vol. 24, 2022.

[3] G. Alexakis, S. Panagiotakis, A. Fragkakis, E. Markakis, and K. Vassilakis, "Control of Smart Home Operations Using Natural Language Processing, Voice Recognition and IoT Technologies in a Multi-Tier Architecture," Designs, vol. 3, pp. 32, 2022.

[4] P. J. Rani, J. Bakthakumar, B. P. Kumaar, U. P. Kumaar, and S. Kumar, "Voice controlled home automation system using Natural Language Processing (NLP) and Internet of Things (IoT)," in 2022 Third International Conference on Science Technology Engineering & Management (ICONSTEM), 2022.

[5] N. Jangamreddy, V. Sethi, and S. Pal, "Web-based gesture recognition system for controlling heterogeneous IoT devices using deep learning," in Proc. 2022 11th Int. Conf. on Communication Systems & Networks (COMSNETS), Bengaluru, India, 2022.