SSRG International Journal of Computer Science and Engineering
ISSN: 2348–8387 / https://doi.org/10.14445/23488387/IJCSE-V11I6P107

Volume 11 Issue 6, 47-52, June 2024

Original Article

Face Recognition-Based Authenticated Voice Assistant System

M. Tirumal[1], M. Sai Teja[1], A. Thushar Babu[1], Kasi Bandla[1*]

[1]*Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India.*

[1]*Corresponding Author : kasi.b@sreenidhi.edu.in*

Received: 23 April 2024 Revised: 28 May 2024 Accepted: 11 June 2024 Published: 30 June 2024

Abstract *- This paper presents an authenticated Voice Assistant System based on Face Recognition. It describes the design and development of a cutting-edge system intended to provide users with a safe and customized interaction. Incorporating a facial recognition authentication component enables the voice assistant to securely identify authorized users before granting access to the system to its comprehensive range of capabilities. This supplementary layer of biometric security ensures that only the rightful user can interact with the system, thereby enhancing security by preventing unauthorized access. Upon successful authentication through facial recognition, the voice assistant will activate a wake-word detection mechanism, facilitating seamless voice-based interaction. This hands-free control feature will empower users to effortlessly access many functionalities beyond basic commands, including application launching, multimedia management, contact handling, web search, and even Chabot integration. The technical considerations for this article involve ensuring the accuracy and reliability of the facial recognition system, implementing secure data storage and access controls, and addressing any potential privacy concerns. These crucial elements will be addressed by the suggested voice assistant to provide a personalized and user-friendly experience that securely and effectively meets the changing needs of contemporary users.*

Keywords *- Voice assistant system, Facial recognition, Machine learning, Hot-word detection, Web search, Application automation, and Chatbot.*

1. Introduction

The landscape of human-computer interaction has undergone a profound transformation with the advent of voice assistants, digital entities that emulate human autonomy and cognitive functions [5]. These sophisticated aids, harnessing cutting-edge technologies such as speech synthesis, voice recognition, and natural language processing, have evolved into indispensable components of contemporary digital interfaces, seamlessly integrating into our daily lives [1]. Through the mechanism of speech recognition, voice control facilitates a myriad of communication possibilities between humans and machines, effectively converting analog signals into digital data and enabling fluid interaction [4]. Beyond their primary function of answering inquiries, voice assistants offer a diverse array of capabilities, including playing music, setting reminders, sending messages, providing real-time weather updates, and managing a spectrum of smart home devices, thereby augmenting convenience and efficiency in various spheres of daily life [1]. Spearheaded by initiatives such as AIVA, ongoing efforts aim to push the boundaries of voice assistant technology, aspiring to develop voice-activated personal assistants endowed with advanced functionalities, such as web browsing and seamless social media interaction, further enriching user experiences and expanding the realm of possibilities [10]. The research underscores the critical importance of accuracy and adaptability in the evolution of voice assistant technology, highlighting their potential to not only streamline tasks but also significantly enhance user engagement and satisfaction across diverse applications and contexts [12]. Despite these advancements, a critical research gap persists in the realms of accuracy and adaptability of voice assistants in diverse and dynamic contexts. Current voice assistants often struggle with understanding nuanced human speech, regional accents, and contextual variations, leading to potential user frustration and limited functionality. Moreover, there is a pressing need for these technologies to adapt more intelligently to individual user preferences and contexts to provide a more personalized and effective interaction experience. Addressing these challenges is crucial for the next generation of voice assistants to not only streamline tasks but also significantly enhance user engagement and satisfaction across a broad spectrum of applications.

This study aims to address this gap by investigating innovative approaches to improve the accuracy and adaptability of voice assistant technology. By focusing on advanced machine learning techniques and user-centric

design principles, this research seeks to develop voice assistants that can better understand and respond to the complexities of human communication, thereby fostering a more intuitive and satisfying user experience. The findings will contribute to the broader field of human-computer interaction by providing insights into the development of more intelligent and responsive digital assistants.

2. Background

In the contemporary era, we are witnessing a shift towards training machines to emulate human thought processes and autonomously perform tasks previously done by humans [1]. This trend has led to the emergence of voice assistants, which are designed to execute various tasks based on spoken commands from users. These voice commands are effectively filtered by the virtual assistant to provide relevant information [2]. This evolution of voice assistant technology has reshaped human-computer interaction paradigms, reflecting a broader trend towards machines emulating human-like cognition and autonomy [3]. These voice assistants, powered by technologies such as voice recognition, speech synthesis, and Natural Language Processing [NLP], have become integral to modern digital experiences [4]. Across the globe, individuals are embracing emerging technologies such as virtual reality, augmented reality, and voice interaction, revolutionizing their digital experiences [5]. Voice control represents a new frontier in human-machine interaction, where analog signals are converted into digital waves through speech recognition [6]. The widespread adoption of smartphones in recent years has led to the increased use of voice assistants like Apple's Siri, Google's Assistant, Microsoft's Cortana, and Amazon's Alexa [7][13]. These assistants leverage technologies such as voice recognition, speech synthesis, and Natural Language Processing [NLP] to offer a wide range of applications, enhancing user convenience and comfort [8]. The concept of voice assistants stems from the need to streamline tasks traditionally performed by humans, with machines now capable of interpreting and executing commands based on voice input [9]. This shift highlights the growing reliance on technology to augment and, in some cases, replace human capabilities in various domains. Voice assistants offer various services to users, including answering questions, playing music, setting timers or alarms, sending messages, providing weather information, and controlling smart devices [10].

Furthermore, the capabilities of voice assistants continue to evolve to meet users' evolving needs [11]. According to Deepak Shende and Ria Umabiya, the AIVA [Artificially Intelligent Virtual Assistant] aims to develop a voice-controlled personal assistant capable of performing various tasks, including internet searches and posting comments on social media platforms, with simple voice commands [3]. Additionally, Tulshan highlighted the potential risks of continuous typing, suggesting adopting voice command systems to mitigate finger injuries [12]. By recognizing and synthesizing spoken words, these systems can efficiently execute tasks. Dr. Kshama V. Kulhalli conducted a survey comparing top voice assistants like Google Assistant, Apple's Siri, and Microsoft's Cortana, concluding that Google Assistant provides the most accurate responses and demonstrates a high level of understanding of voice variations [5]. This underscores the importance of accuracy and adaptability in voice assistant technology to enhance user experiences effectively.

While voice assistants have revolutionized human-computer interaction, they can be susceptible to background noise or misinterpreting accents. To address these limitations and enhance security, our proposed system incorporates facial recognition technology alongside voice interaction. This multimodal approach provides a more robust and secure user experience. Users first verify their identity through a facial scan. Upon successful recognition, the system grants access to the voice assistant's functionalities, allowing for interaction through voice commands or text input. This integration strengthens user privacy and ensures that only authorized users can leverage the system's capabilities.

Our proposed system builds upon the foundation of voice assistants by incorporating facial recognition technology. This additional layer of security ensures that only authorized users can access the voice assistant's functionalities. Upon launching the application, the user's face is scanned and compared against a pre-enrolled database. Successful recognition grants access to the voice assistant interface, allowing for interaction through voice commands or text input. This integration enhances user privacy and strengthens the overall security of the system.

3. Proposed System

This section presents our proposed scheme and also algorithms required for the realization of the scheme.

3.1. Problem Definition

The challenge lies in creating a voice assistant system that incorporates facial recognition for user authentication while ensuring seamless integration with voice interaction. This requires balancing user convenience and security to provide broad functionality effectively. The challenge lies in creating a voice assistant system that incorporates facial recognition for user authentication while ensuring seamless integration with voice interaction. This requires balancing user convenience and security to provide broad functionality effectively.

3.2. Face Recognition Integrated with Voice Assistant

The proposed voice assistant system with facial recognition for secure user authentication. Once a user is

verified, voice commands can launch applications (WhatsApp, YouTube), manage contacts (send messages and calls), control media, and perform web searches. Facial recognition adds security, while voice interaction offers a convenient, hands-free experience. Future advancements can refine accuracy and expand functionality for a more user-friendly and secure voice assistant. The advantages of the proposed system include Facial recognition adds an extra layer of security compared to just voice wake words, making unauthorized access more difficult. Convenient Interaction: Voice commands allow for hands-free control, improving user experience. Broad Functionality: The system goes beyond basic commands, letting you launch apps, manage contacts, control media, and perform web searches through voice.

4. Methodology

The materials used here are described with the methodology.

4.1. Tools Used

During the development of the voice assistant, we utilize tools such as Python and SQLite for backend functionalities, while employing HTML, CSS, and JavaScript for frontend development. Additionally, we employ VS Code as our Integrated Development Environment (IDE).

4.1.1. Libraries Installed in Python

Python has various libraries, which include OS, pipes, re, struct, subprocess, webbrowser, play-sound, pyaudio, pyautogui, pywhatkit, multiprocessing, SQLite3, and CSV. These libraries allow for operating system-dependent functionality, regular expression support, and conversion between Python values and C structs. They also provide high-level interfaces for web-based documents, audio playback, face recognition, and multiprocessing, as well as utilities for working with WhatsApp and SQLite databases.

4.2. Methodology

The process starts with the "Start" phase, leading to "Face Detection" which implies the system's use of a camera or visual input to identify the user's face, subsequently branching to "Facial Data" for speaker identification through facial feature recognition. Upon a successful match, the system switches to an open "Web Application" for text-to-speech functionality, directing to a "User Interface" section offering text input and control options. This segment is further divided into "Hot-word Detection" and "Text Input". Upon detecting a predefined "Hot-word", the system activates "Speech Recognition" to capture spoken input. Both paths then converge into an "All Interfacing" section, likely integrating various input sources. From there, "Output Tend Result" prepares processed data for the subsequent stage, leading to "Text to Speech Conversion", ultimately generating spoken audio output.

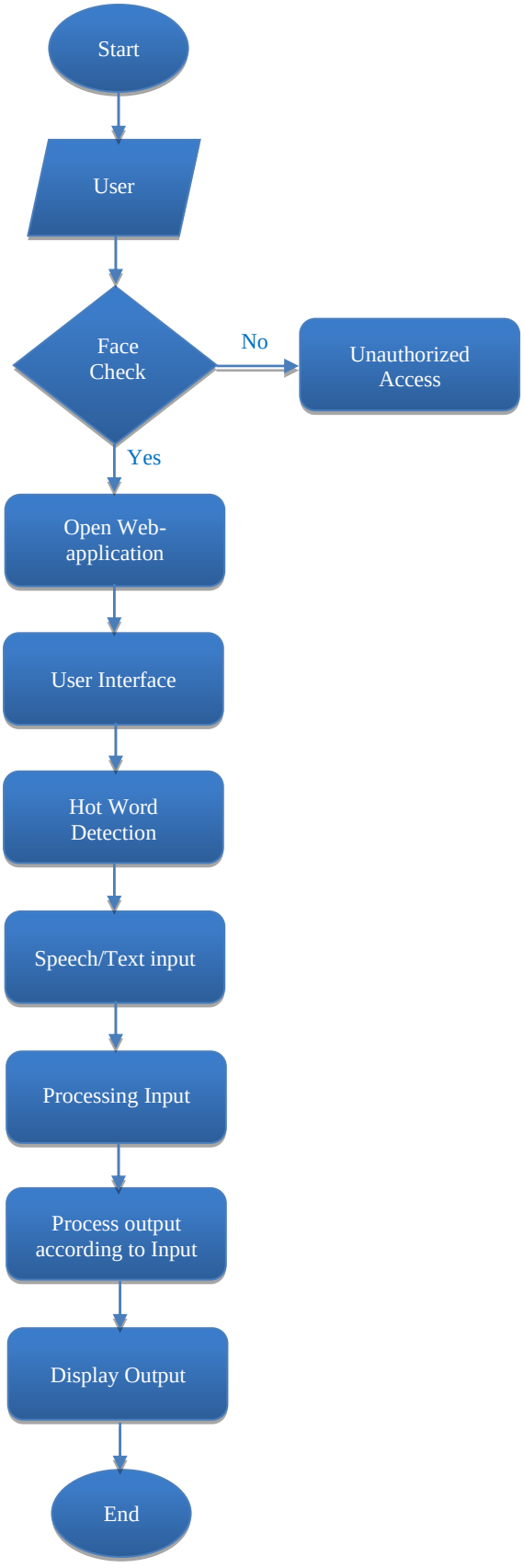

Fig. 1 Methodology flowchart

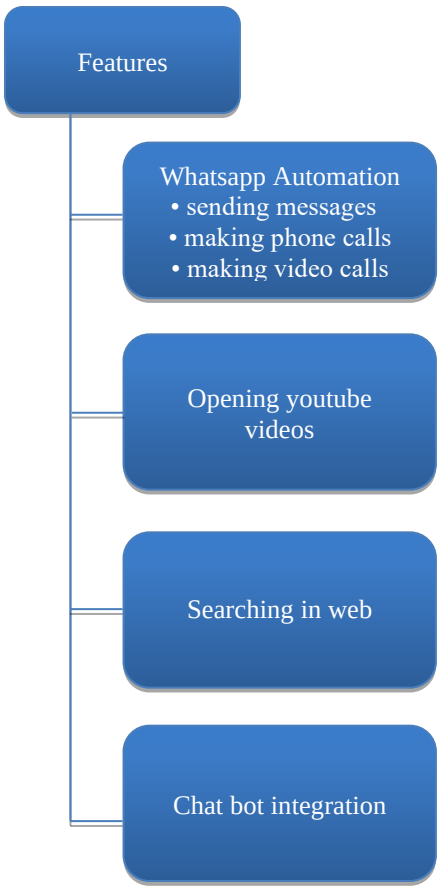

Fig. 2 Features of voice assistant

The system offers various features for users, including application management, contact management, media control, web search, and optional Chabot integration. Users can launch installed applications by simply speaking their names or providing their paths, accessing a pre-defined database or using stored URLs. The assistant integrates with your contact database, allowing users to search for phone numbers and initiate actions like sending WhatsApp messages or making calls. Voice commands enable easy control of media playback, such as playing YouTube videos or adjusting audio settings. The assistant can perform web searches using your preferred search engine, such as Google, and can provide more interactive and conversational experiences. The system may also offer future integration with Chabot libraries for enhanced interaction.

4.3. Features

The features begin with facial recognition integration, enabling face detection and subsequent speaker identification. The system facilitates application launching by mapping user commands to respective application paths or URLs. Additionally, it enables YouTube video playback based on user queries and incorporates hot word detection functionality using the Porcupine library. Contact search and interaction capabilities are provided, allowing users to initiate calls or messages via WhatsApp. Users can also perform Google searches directly through the system. Integration with a chatbot enables conversational interaction, and audio playback functionality is included. Utilizing the Eel library, the system offers a user interface accessible through a web application. Furthermore, robust mechanisms ensure smooth operation and a user-friendly experience throughout interactions.

5. Results and Discussion

5.1. Results

The research paper discusses the development of a Python-based voice assistant system, emphasizing its key features, such as facial recognition integration, application launching, and third-party library integration for tasks like YouTube playback and Chabot interaction. It also highlights the importance of error-handling mechanisms for ensuring system reliability. Overall, the paper showcases the system's potential to enhance user interaction and productivity through advanced functionalities.

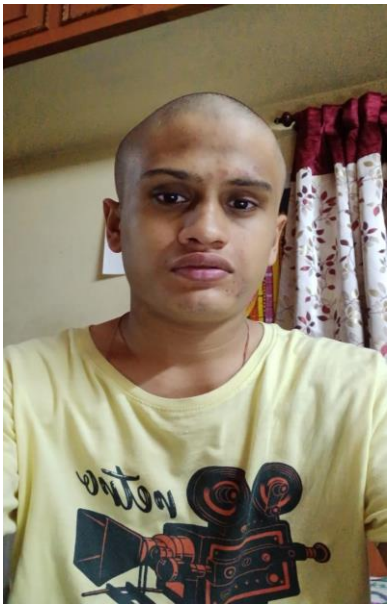

Fig. 3 Face recognition

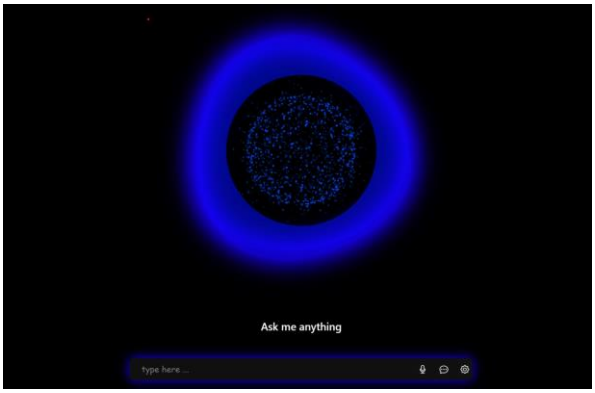

Fig. 4 Voice assistant display

Fig. 5 YouTube and web search automation

A secure voice assistant unlocks upon successful facial recognition, enabling hands-free interaction through voice commands or text input shown in Figure 4.

The YouTube Automation result image demonstrates the efficacy of an algorithm in automating video retrieval and optimizing search accuracy and efficiency. Similarly, the Web Search Automation result image showcases an automated system that enhances web search processes, improving the precision and retrieval of pertinent information. Both images highlight significant advancements in automation technology depicted in Figure 5.

Figure 6 shows the Chat feature, which helps the voice assistant to chat.

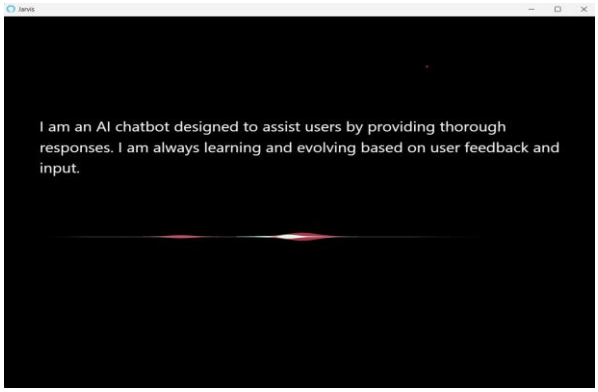

Fig. 6 Chat feature

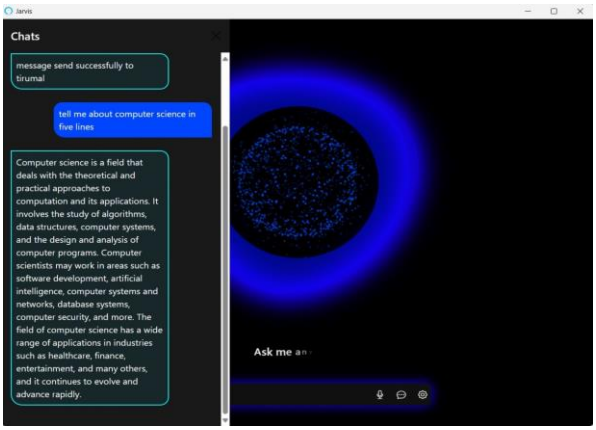

Fig. 7 Chat history

Figure 7 shows the chat history, which helps us to know all the actions performed using the voice assistant.

5.2. Limitations

The proposed system is designed to manage standard tasks efficiently, but they may face limitations when dealing with heavier workloads or complex operations. These include performance bottlenecks, increased processing time, and resource utilization.

Heavy usage can also reduce responsiveness and system slowdowns, impacting user experience. Therefore, careful evaluation and enhancements may be necessary for optimal performance.

6. Conclusion and Future Scope

The suggested facial recognition voice assistant system provides a safe and extensive feature set. The system can handle a variety of user demands through natural language processing and action execution, and facial recognition improves security over voice-only authentication. Numerous applications have become possible as a result, including managing contacts, operating smart home appliances, and doing web searches. To address user privacy concerns and increase facial recognition security and accuracy, ongoing development is essential. Examining user customization choices can also improve the user experience. Voice assistants with facial recognition have the potential to develop into a potent and seamless means of interacting with technology when these issues are resolved as the field advances.

This proposed voice assistant system distinguishes itself from existing systems through its advanced authentication mechanism that utilizes facial recognition. Traditional voice assistant technologies typically rely on voice commands alone for activation, which can be susceptible to unauthorized access and misuse. Our system, however, introduces a significant enhancement by incorporating facial recognition, ensuring that only authorized users can activate and interact with the assistant. This dual-layered security approach not only mitigates the risks associated with unauthorized access but also enhances the overall user experience by providing a seamless and secure interaction. This innovation addresses a critical gap in current technologies, offering a more robust and reliable solution for secure voice assistance.

References

[1] Yuqi Huang, "Research on the Development of Voice Assistants in the Era of Artificial Intelligence," *2022 2nd International Conference on Social Development and Media Communication (SDMC 2022)*, Sanya, Hainan, China, vol. 115, pp. 1-4, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ajay Sahu et al., "Voice Assistant Using Artificial Intelligence," *Proceedings of the International Conference on Innovative Computing and Communication (ICICC)*, pp. 1-4, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] G. Preethi et al., "Voice Assistant using Artificial Intelligence," *International Journal of Engineering Research and Technology (IJERT)*, vol. 11, no. 5, pp. 453-457, 2022. [Publisher Link]

[4] Anthony Zhang, SpeechRecognition 3.10.4, 2022. Online. [Available]: https://pypi.org/project/SpeechRecognition/

[5] Geekpradd, PyDictionary 2.0.1, 2020. Online. [Available]: https://pypi.org/project/PyDictionary/

[6] AlSweigart, PyAutoGUI 0.9.54, 2023. Online. [Available]: https://pypi.org/project/PyAutoGUI/

[7] Natesh M Bhat, pyttsx3 2.90, 2020. Online. [Available]: https://pypi.org/project/pyttsx3/

[8] M. Nallamothu, and R. Mukkamala, "A Study on Speech Recognition and Desktop Assistant using Python," *Journal of Emerging Technologies and Innovative Research*, vol. 6, no. 4, pp. 521-526, 2019.

[9] Amrita S. Tulshan, and Sudhir Namdeorao Dhage. "Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa," *Advances in Signal Processing and Intelligent Recognition Systems: 4th International Symposium SIRS 2018*, Bangalore, India, pp. 190–201, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] A. Haseeb, and S.A.R Rizvi, "A Review of Voice Assistant and Its Impact on the Healthcare Industry," *International Research Journal of Engineering and Technology*, 2022.

[11] John Rampton, Advantages and Disadvantages of Voice Assistants for Marketers, Oracle Marketing Cloud Blog, 2022. Online. [Available]: https://blogs.oracle.com/marketingcloud/post/advantages-and-disadvantages-of-voice-assistants-for-marketers.

[12] Shailesh Shyamlal Gupta, "Artificial Intelligence & its Applications," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 5, no. 10, pp. 1857-1861, 2023. [Publisher Link]