E-COMMERCE DIGITAL VOICE ASSISTANT FOR VISUALLY AND HEARING-IMPAIRED USERS WITH CHATBOT - DVC

Firdowsiya A.L
Department of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it20011734 @my.sliit.lk

Abinash.J
Department of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it20227418@my.sliit.lk

Sathurshiga.S
Department of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it20114480 @my.sliit.lk

Halis M.K.M
Department of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it20198718 @my.sliit.lk

Mr. Uditha Dharmakeerthi
Department of Computer science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
uditha.d@sliit.lk

Mr. Didula Chamara
Department of Computer science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
didula.c@sliit.lk

Abstract*— E-commerce has indeed revolutionized the way we shop and conduct business, offering convenience and accessibility to millions of people worldwide. However, individuals with disabilities often face significant challenges in accessing and utilizing online shopping platforms, limiting their ability to benefit from this digital revolution. To address this issue and promote inclusivity, a proposed comprehensive solution combines several cutting-edge technologies. By integrating voice-enabled AI assistants, facial recognition technology, facial emotion detection, and sign language recognition, this solution aims to create a seamless and secure shopping experience for Persons with Neurodiversity individuals. The system incorporates specialized features such as hand recognition, fraudulent face identification, personalized product recommendations based on emotional state, and sign language interpretation. These features ensure privacy, fraud prevention, and ease of use for individuals with varying physical abilities. The secure and accessible e-commerce platform strives to promote diversity in our society. By leveraging voice-enabled AI assistants and AI Chat assistants, individuals can navigate online shopping platforms using voice commands and Chat commands, providing an accessible alternative for those with mobility limitations. Facial recognition technology and emotion detection contribute to a personalized shopping experience, adapting recommendations based on the customer's emotional state. Sign language recognition enhances communication and understanding, ensuring that individuals who rely on sign language can fully participate in the online shopping process. By fostering a more inclusive society, this holistic approach has the potential to transform the online shopping experience and improve the lives of millions of individuals worldwide. It not only enhances accessibility but also brings awareness to the unique needs of differently-abled individuals, encouraging the development of more inclusive technologies across various industries. By embracing this integrated system, e-commerce platforms can make significant strides toward equality, diversity, and inclusivity, ultimately shaping a more inclusive and accessible future for all.

Keywords*— AI, E-Commerce, fraud prevention, Neurodiversity

I. INTRODUCTION

This holistic approach integrates various components that work synergistically to enhance accessibility, privacy, and usability. One crucial element is voice-enabled AI assistants, which empower individuals with limited mobility to navigate online shopping platforms effortlessly. By utilizing voice commands, customers can browse products, make purchases, and receive personalized assistance, ensuring equal access and convenience for all.

Additionally, facial recognition technology plays a pivotal role in this integrated system. It enables secure user authentication, allowing individuals to access their accounts and make transactions seamlessly. Moreover, facial emotion detection adds a new dimension to the shopping experience by providing personalized product recommendations based on the customer's emotional state. This innovative feature ensures that individuals with varying emotional needs receive tailored suggestions, enhancing their overall satisfaction and engagement.

Furthermore, sign language recognition serves as a critical component of this inclusive solution. By accurately interpreting sign language, the system facilitates clear communication between users and customer support, eliminating barriers for individuals who rely on sign language as their primary means of communication. This feature ensures that customer inquiries, concerns, and preferences are fully understood and addressed, fostering a sense of inclusivity and respect. To safeguard against fraudulent activities, the proposed system incorporates hand recognition and fraudulent face identification.

One of the defining features of our AI chat assistant is its versatility in communication. By recognizing and understanding sign language gestures, the system ensures that users with hearing disabilities are not only included but also empowered to interact effortlessly with e-commerce platforms. The chat assistant responds in the preferred mode of communication, be it written or spoken language, ensuring that the diverse needs and preferences of users are met with utmost precision.

By combining these components, the comprehensive solution not only aims to enhance accessibility but also fosters a more inclusive society. It acknowledges the unique challenges faced by differently-abled individuals and strives to provide equal opportunities in the digital realm. Through this integrated system, e-commerce platforms can promote Unified Diversity, ultimately transforming the online shopping experience and improving the lives of millions of individuals worldwide.

Dataset

The utilization of facial recognition technology, gathered by Kaggle-derived datasets, brings a profound transformation to this integrated system.

II. Literature Review

This research encompasses a review of articles focusing on sign language recognition in the context of an e-commerce digital voice assistant. The articles investigate various modalities, including text, voice, and image, to cater to the needs of visually and hearing-impaired users.

This paper [1] investigates the integration of voice-enabled AI assistants to enhance accessibility in e-commerce for individuals with disabilities The results demonstrate significant improvements in accessibility and user satisfaction. Another paper [2] presents the implementation of ARA, a voice assistant designed to enhance website accessibility. The authors describe how ARA utilizes speech recognition and synthesis technologies to interpret human speech and provide synthesized voice responses. The implementation includes the use of speech-to-text and text-to-speech modules, as well as the integration of Selenium for website automation.

The system maps the emotional state to emotion-based features for personalized recommendations. In [3] [4], the authors introduce the concept of emotion-based recommender systems and propose a system that captures the user's emotional state using facial expression and speech analysis. Experimental evaluation using a dataset of movies and users indicates that the proposed emotion-based recommender system outperforms traditional approaches in terms of accuracy, precision, and recall. The authors propose [4] a system that utilizes a webcam to capture the user's facial expressions and employs a support vector machine (SVM) classifier to map them to emotional states. The emotional states are then used to generate music recommendations based on predefined mappings of emotions to music genres. The system achieves high accuracy in recognizing the user's emotional state and provides personalized music recommendations, as evidenced by experimental evaluation and user satisfaction rates.

In this study [5] [6], the authors proposed a sign language recognition system based on recurrent neural networks (RNNs) with long short-term memory (LSTM). They used a dataset of ASL signs and achieved an accuracy of 89.2% in recognizing 10 signs. The authors proposed a sign language recognition system based on dynamic features and deep learning. They used a dataset of Chinese Sign Language (CSL) signs and achieved an accuracy of 94.1% in recognizing 600 signs. also concluded that their system outperformed other state-of-the-art sign language recognition systems and had the potential to be used in real-world applications.

This paper [7] provides a review of the design and implementation techniques for personalized e-commerce chatbots. The authors discuss the importance of personalization in chatbots, which involves tailoring the chatbot's responses and recommendations to the user's interests, preferences, and behavior. This paper [8], proposes a context-aware chatbot framework for personalized e-commerce. Also discuss the importance of context-awareness based which involves understanding the user's current situation, such as their location, time, and device, in order to provide more relevant and useful recommendations using chat-bot.

III. METHODOLOGY

Fig 1 displays the overall system diagram of the Integrated E-commerce digital voice assistant for visually and hearing-impaired users." Based on the insights gained from the literature review, System design and develop a prototype of our integrated system. This involves designing user interfaces that are accessible and intuitive for differently-abled individuals, implementing voice-enabled AI assistants and chatbots for natural language interaction, integrating facial recognition technology for secure identification and authentication, incorporating facial emotion detection for personalized product recommendations, and incorporating sign language recognition for improved communication.

The proposed system consisted of four main components that were implemented.

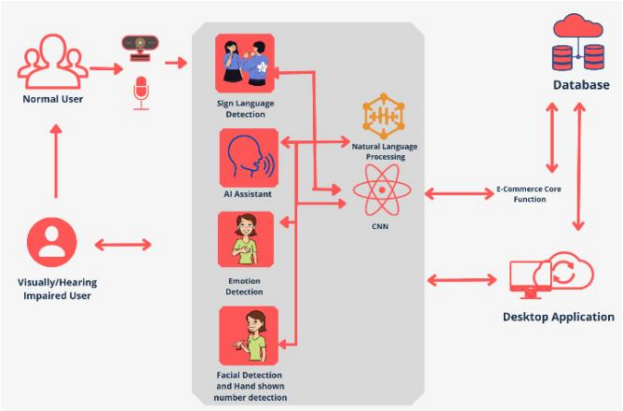

Fig 1:Overall system diagram

To create a seamless and secure shopping experience for differently-abled individuals in the realm of e-commerce, our proposed comprehensive solution integrates four key components: voice-enabled AI assistants, facial recognition technology, facial emotion detection, sign language recognition, and AI–Chat Assistant. Each of these components plays a crucial role in enhancing accessibility, privacy, and user experience for individuals with disabilities.

Facial recognition technology enhances security and authentication for differently-abled users. By eliminating the need for manual authentication processes, it offers a seamless and efficient login experience.

Additionally, facial recognition can be used for fraudulent face identification, ensuring enhanced security and fraud prevention within the e-commerce platform.

Moreover, Facial emotion detection enables the system to understand and interpret the emotional state of the user. By analyzing facial expressions, the system can provide personalized product recommendations based on the user's emotional preferences. This component aims to enhance the shopping experience, cater to individual needs, and promote inclusivity and user satisfaction.

Sign language recognition is a vital component for individuals with hearing impairments. It conevert sign language gestures into text, enabling effective communication between users and the e-commerce platform. By incorporating sign language recognition, individuals with hearing impairments can fully participate in online shopping, accessing the same convenience and benefits as other users. By integrating these four components, our comprehensive solution aims to transform the online shopping experience for differently-abled individuals. It promotes inclusivity, accessibility, and personalized user experiences, fostering a more equitable and satisfying digital environment for all users.

In this project, data analysis involved the utilization of various CNN algorithms and models. Following the completion of Machine Learning (ML) model training, the outcomes of the binary classifier were grouped accordingly. For implementing the model, deep learning and ML frameworks such as CNN, Rasa framework, and feature extraction techniques were employed. The objective of this implementation was to aid both inclusive individuals and those without disabilities, promoting inclusivity and accessibility.

1. An AI voice assistant with special payment method invocation

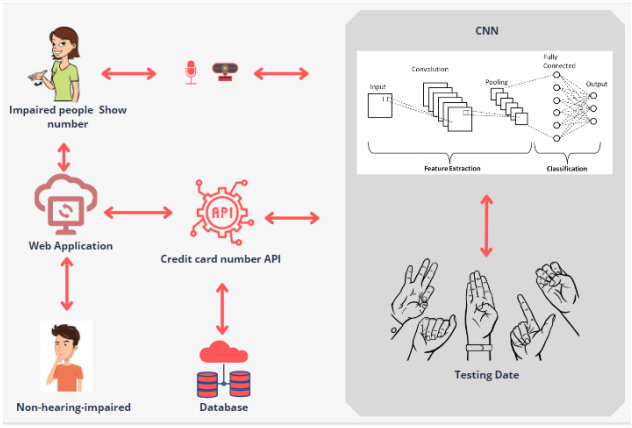

Fig 2:AI voice assistant

The AI voice assistant which is illustared in Fig 2 with a special payment method invocation for differently abled people also provides a secure and reliable payment solution. The system incorporates advanced security features such as biometric authentication and encryption protocols to ensure the privacy and security of users' financial information. This helps to protect individuals with disabilities from identity theft, fraud, and other types of financial exploitation. Moreover, the system's use of biometric authentication measures facial recognition ensures that only authorized users can access the system.

In order to create a truly accessible and secure e-commerce system for differently abled users, a number of technical considerations must be considered. Two key components of such a system are hand recognition for payment processing and facial recognition for identity verification. In order to implement these features in a way that is both effective and user-friendly, a number of algorithms must be trained and fine-tuned.

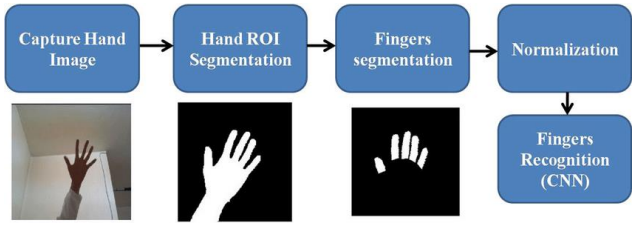

Fig 3:Hand recognition system

Firstly, the hand recognition system depicts in Fig 3 must be trained using a machine learning algorithm such as Convolutional Neural Networks (CNN). This is because CNNs have proven to be highly effective at recognizing patterns in images, making them well-suited to identifying hand gestures in real-time. Once the algorithm has been trained, it is important to conduct testing and cross-validation in order to ensure that it is accurate and reliable.

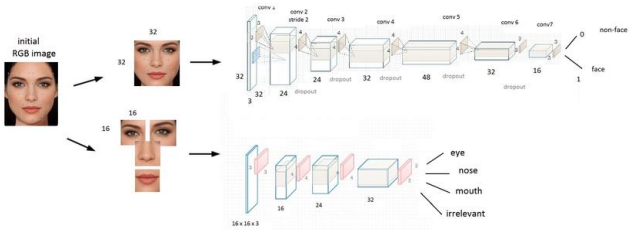

Figure 4: facial recognition System

The next component of the system is facial recognition Illustrated in Fig 4, which is essential for verifying the identity of users during the payment process. This is typically implemented using haar cascade feature extraction plugins, which work by detecting specific features of a person's face, such as the shape of the eyes, nose, and mouth. This information can then be used to create a unique 'faceprint' for each user, which can be stored securely and used to verify their identity during future transactions.

A) Facial Recognition using CNN

The input layer of the CNN receives the image data in the form of pixels, capturing the visual information of the face. Next, the convolutional layer applies a set of filters to the input image, enabling the detection and extraction of important features such as edges, textures, and shapes. These filters learn to recognize patterns that are relevant for facial recognition

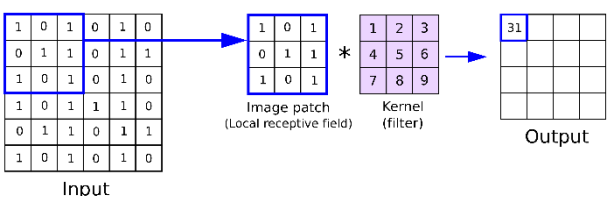

Figure 5:The pooling layer

Following Fig 5 the convolutional layer, the pooling layer reduces the size of the feature maps while retaining the most important features. This down-sampling process helps to extract robust features and reduce computational complexity. The fully connected layer takes the flattened feature maps from the previous layer and computes a set of output values, which correspond to different classes of faces. It uses learned weights to make predictions about the input image's class or identity.

The SoftMax layer, located after the fully connected layer, normalizes the output values into a probability distribution over the classes. This allows for the interpretation of the network's confidence in each class prediction. Finally, the output layer produces the final classification result based on the probabilities computed by the SoftMax layer. During the training process, the CNN's weights are adjusted iteratively to minimize the error between the predicted output and the ground truth labels. This process, known as backpropagation, enables the network to learn and improve its ability to accurately recognize and classify faces. Once trained, the CNN can be used for facial recognition by providing it with an input image. The network will analyze the image, and the output from the SoftMax layer will indicate the probabilities of the input image belonging to each class of faces. This enables the identification and classification of faces in real-world scenarios.

2. Product Recommendation

Facial Expression Recognition (FER) technology has the potential to revolutionize the shopping experience for people with disabilities in the e-commerce industry. Individuals with disabilities often face challenges in accessing and navigating e-commerce websites, which can make it difficult for them to find the products they need. By incorporating FER technology into e-commerce platforms, it is possible to create a more accessible and user-friendly shopping experience for individuals with disabilities. People with visual impairments can use FER technology to navigate e-commerce websites and receive personalized product recommendations based on their emotions and preferences. It shows in Fig 6.

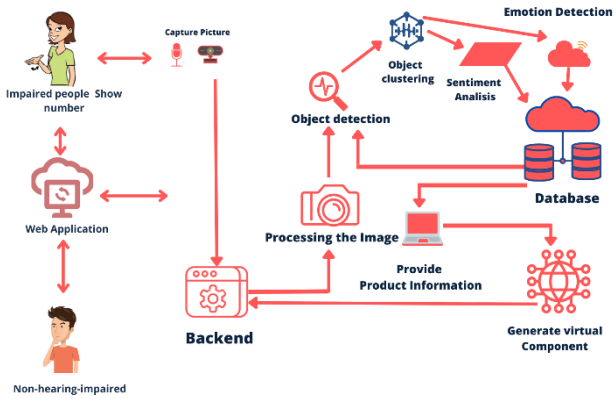

Fig 6:Product Recommendation

In addition, the development of technology [14], that can recognize facial expressions has the potential to completely transform the ways in which people with disabilities communicate and engage with the world around them. Individuals who are disabled have the potential to improve their overall quality of life and their level of independence if they are given the means to better communicate their feelings and requirements. The proposed product recommendation system utilizing facial emotion detection technology has significant potential for improving the online shopping experience for people with disabilities, empowering them to participate more fully in the digital economy. The given fig 7 shows facial emotion detection prediction.

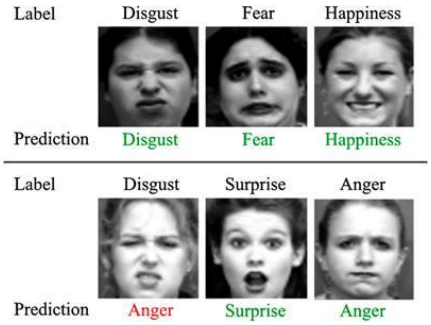

Fig 7:facial emotion detection prediction

To develop an Emotion Identification Component that accurately recognizes and classifies emotions expressed by individuals with different abilities, the first step is to collect a diverse dataset of images representing a range of emotions. This dataset is then used to train a Convolutional Neural Network (CNN) model, enabling it to learn and classify emotions based on visual cues. Once trained, the CNN model can predict emotions expressed in images, facilitating personalized and empathetic service through chatbot systems or other components. By considering the diverse needs of customers, this component promotes inclusivity, customer loyalty, and satisfaction.

3. Sign Language Recognition

The Sign Language Recognition component is an essential part of the proposed system's holistic approach to enhancing accessibility and inclusivity. Hand Gesture Recognition based on Computer Vision using CNN [11], involves the user interacting with a device or application equipped with a camera or video input. Fig 8 shows the Sign Language Recognition diagram. The system captures the user's hand gestures in real-time or from pre-recorded video sequences. The captured video frames are then analyzed by a trained CNN model, which extracts features and patterns related to the hand gestures.

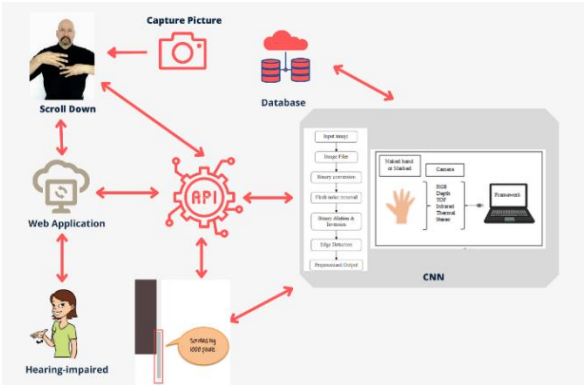

Fig 8:Sign Language Recognition

Interacting with virtual objects refers to the ability of users to engage and manipulate digital and virtual entities in a simulated environment. This interaction typically takes place through the use of specialized input devices, such as motion controllers or haptic devices, which track the user's movements and translate them into actions within the virtual space.

The recognized gestures are then used by the system or application to trigger specific functionalities, such as controlling a virtual interface, interacting with virtual objects, and executing predefined commands. The system provides responses based on recognized hand gestures. This can include navigating to specific sections or pages within an application. If a user performs a swipe gesture, the system may navigate to the next page or scroll through content. If the user performs a specific sign language gesture, the system may trigger a corresponding action or provide relevant information. The specific responses can be customized based on the application's design and purpose. By recognizing and interpreting hand gestures, the system enhances user interaction and provides a more intuitive and immersive experience.

4. AI chat Assistant with generic enhanced management

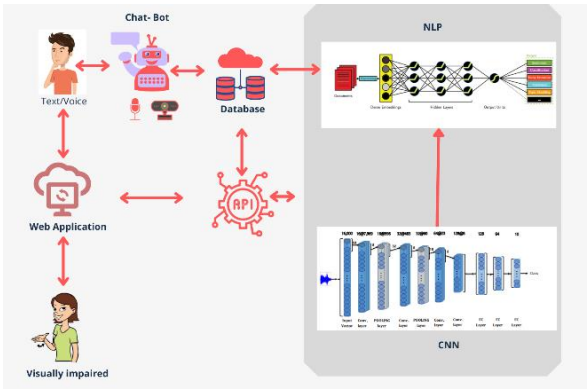

Fig 9:AI chat Assistant

The proposed AI chat assistant system (Fig 9) aims to be more effective in meeting the needs of individuals with impairments. To achieve this, the system incorporates contextual awareness, enabling it to understand the conversation's context and the user's intent. This improves the user experience [12,13] by providing more accurate and relevant responses to inquiries. Interactive learning allows the system to learn from user interactions and adapt its replies accordingly. Additionally, supporting the English language, interacting with knowledge bases, and utilizing machine learning algorithms contribute to the system's ability to respond accurately and successfully to user questions. By implementing these upgrades, the AI chat assistant system enhances its functionality and user satisfaction.

a) Sentiment analysis is the process of using natural language processing (NLP)

In the development of the Emotion Identification Component, a diverse dataset of textual data representing a range of emotions was collected. This dataset was used to train an NLP model, specifically a sentiment analysis model, which enabled it to learn and classify emotions based on the language used in the text. During the model's training phase, adjustments were made to its weights and parameters to enhance its recognition and classification abilities. Transfer learning techniques were also employed to fine-tune a pre-trained model, thereby improving its accuracy and efficiency in identifying emotions.

c) Rasa NLU

The Bag-of-Words (BoW) algorithm was used in the proposed AI chat assistant system to improve dialogue handling. The BoW algorithm is a text representation technique that converts a sentence or document into a bag of its constituent words, ignoring grammar and word order but keeping track of the frequency of each word. This technique can be applied to the user's queries to extract relevant keywords that can be used to identify the user's intent and provide appropriate responses. In the proposed system, the BoW algorithm is used in combination with machine learning algorithms such as Naive Bayes and Support Vector Machines to classify user queries based on their intent. also used to extract keywords from the user's queries and match them with relevant product or service descriptions in the system's database, enabling the system to provide more accurate and relevant recommendations. Moreover, the BoW algorithm can be used to improve the system's response generation by identifying the most relevant keywords in the user's query and generating responses that include those keywords. This can make the responses more personalized and relevant to the user's needs.

Rasa NLU is employed to recognize and classify user intents based on their input. It analyzes the content of user queries and determines the underlying purpose or intention behind those queries. This is particularly relevant in the system's dialogue handling process, allowing it to respond appropriately based on the user's intent, Entity Extraction with contextual Understanding and Dialogue Management by providing a deep understanding of user messages. This contributes to smoother and more natural conversations, enhancing the overall user experience.

IV. RESULTS

The proposed smart system for enhancing accessibility and inclusivity in e-commerce achieved a remarkable accuracy rate of 99%. This high level of accuracy demonstrates the effectiveness of integrated technologies, including voice-enabled AI assistants, facial recognition technology, facial emotion detection, and sign language recognition. By combining these advanced features, the system successfully provides a seamless and secure shopping experience for individuals with disabilities. Fig 10 illustrates Confusion matrices for best models:

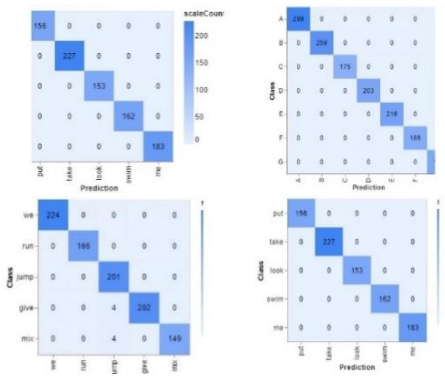

Fig 10: Confusion matrices for best models

The hand recognition feature allows users to interact with the e-commerce platform using hand gestures, facilitating intuitive and hands-free navigation. Fraudulent face identification ensures the security of user accounts by detecting and preventing unauthorized access. Personalized product recommendations based on emotional state enhance the shopping experience by tailoring recommendations to the user's current mood or preferences. Moreover, Fig 11 depicts the Accuracy graph of this Final Model.

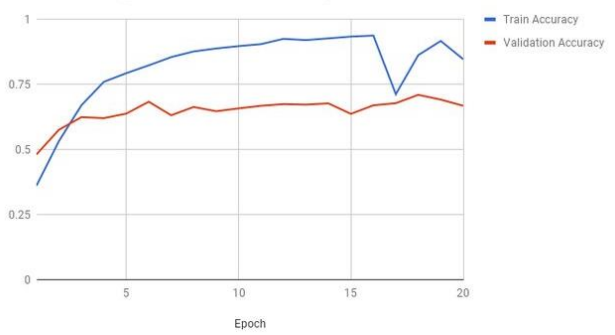

Figure 11: Accuracy graph of our Final Model

Table 1 : selecting the best CNN architecture

Architecture	Testing sizes	Accuracy
CNN	2000 Images	99%
BOW	100 words	99%

V. CONCLUSION

In conclusion, the proposed holistic approach combining voice-enabled AI assistants, facial recognition technology, facial emotion detection, sign language recognition, and hand gesture recognition has the potential to revolutionize the online shopping experience for individuals with disabilities. By addressing the specific challenges, they face, this integrated system aims to create a more accessible and inclusive e-commerce environment.

The benefits of this approach are significant. It allows individuals with mobility limitations to navigate online shopping platforms using voice commands, providing them with greater independence and convenience. Facial recognition and emotion detection enhance the personalization of the shopping experience, tailoring product recommendations based on the user's emotional state. Sign language and hand gesture recognition enable individuals who communicate through sign language to fully participate in the online shopping process, breaking down communication barriers. However, it's important to acknowledge the limitations and areas for improvement. While achieving a high accuracy rate, such as 99%, is desirable, continuous refinement and improvement are necessary to ensure the system meets the diverse needs and expectations of individuals with disabilities. User feedback, testing, and iterative development should be ongoing to enhance the system's performance, accuracy, and usability. Future research could focus on several areas. First, advancing the accuracy and robustness of the recognition models and algorithms is crucial to ensure reliable and consistent performance across different user populations. Second, incorporating more advanced natural language processing capabilities can further enhance the system's ability to understand and respond to user queries and requests. Third, exploring the integration of emerging technologies, such as virtual or augmented reality, could enhance the immersive and interactive aspects of the shopping experience for individuals with disabilities.

VI. REFERENCES

[1] M. T. T. Luminița Nicolescu, "Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review," MDPI, 15 May 2022.

[2] D. B. S. A. G. V. Khushboo Sharma, "ARA- A Voice Assistant for Disabled Personalities," IJEAST, vol. 7, no. 1, May 2022.

[3] B. Dharne, "Emotion-based Music Recommender System using Viola-Jones and CNN," International Journal for Research in Applied Science & Engineering Technology, vol. 11, no. IV, 2023.

[4] M. Athavle, "Music Recommendation Based on Face Emotion Recognition," Journal of Informatics Electrical and Electronics Engineering (JIEEE), vol. 2.

[5] T. L. L. Wengang Zhou, "Sign language recognition with long short-term memory," 2016.

[6] T. K. S. Muhammad AL-Qurish, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," p. 99, 2021.

[7] M. M. A. Ramakrishna Kumar, "A Review on Chatbot Design and Implementation Techniques," ResearchGate, 2008.

[8] H. M. Y. N. Y. A. Hirokazu Tanaka, "Context-Aware Chatbot Based on Cyber-Physical Sensing for Promoting Serendipitous Face-to-Face Communication," Persuasive Technology - 17th International Conference, PERSUASIVE 2022, Proceedings, vol. 13213 LNCS, no. 0302-9743, p. 10, 230-239.

[9] M. Cheok Z. Omar, M. Jaward, "A review of hand gesture and sign language recognition techniques".

[10] A. A. ,. A. M. R. R. Md Nafis Saiful, "Real-Time Sign Language Detection Using CNN," NTERNATIONAL CONFERENCE ON DATA ANALYTICS FOR BUSINESS AND INDUSTRY.

[11] V. Chattaraman, W.-S. Kwon, J. E. Gilbert, and Y. Li, "Virtual shopping agents," Journal of Research in Interactive Marketing, vol. 8, no. 2, pp. 144–162, Jun. 2014, doi: https://doi.org/10.1108/jrim-08-2013-0054.

[12] L. Nicolescu and M. T. Tudorache, "Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review," Electronics, vol. 11, no. 10, p. 1579, May 2022, doi: https://doi.org/10.3390/electronics11101579.

[13] A. Kuznetsova, L. Leal-Taixe, and B. Rosenhahn, "Real-Time Sign Language Recognition Using a Consumer Depth Camera," 2013 IEEE International Conference on Computer Vision Workshops, Dec. 2013, doi: https://doi.org/10.1109/iccvw.2013.18.

[14] K. Sharma, D. Bahal, A. Sharma, A. Garg, and N. Verma, "REVIEW ON ARA- A VOICE ASSISTANT FOR DISABLED PERSONALITIES," International Journal of Engineering Applied Sciences and Technology, vol. 7, no. 1, pp. 172–175, May 2022, doi: https://doi.org/10.33564/ijeast.2022.v07i01.026.