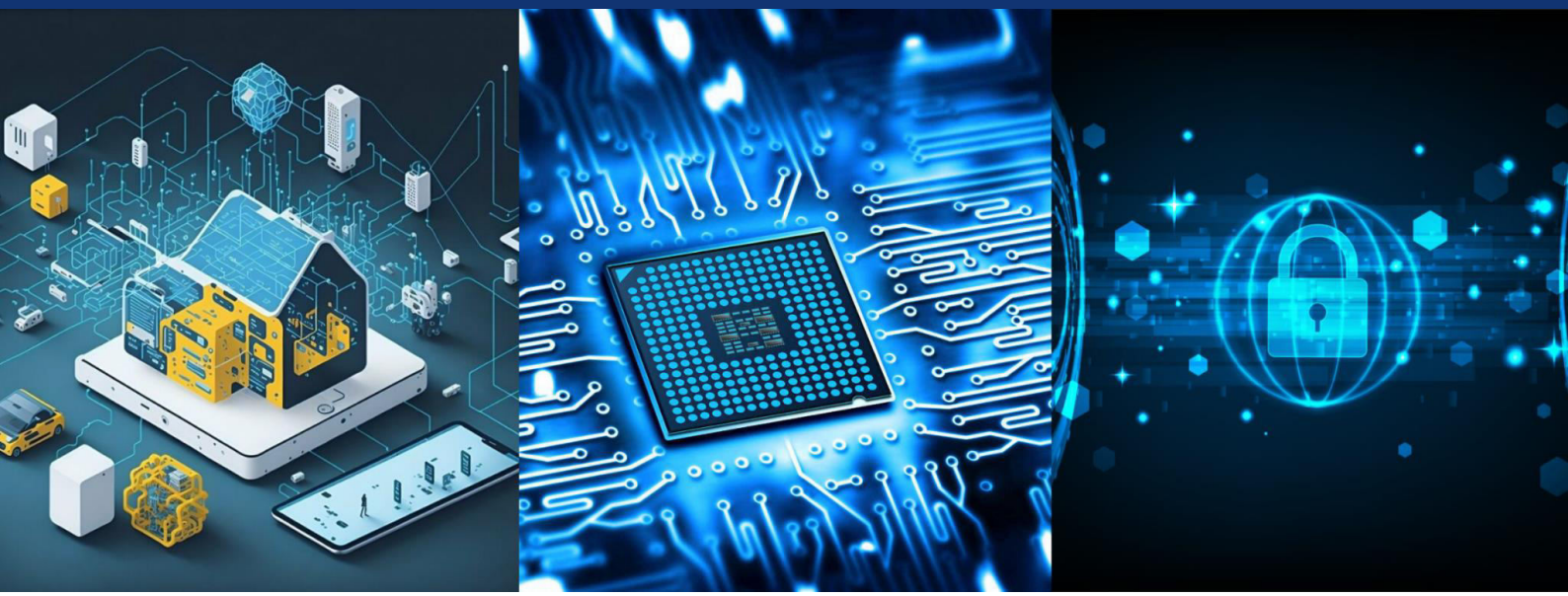
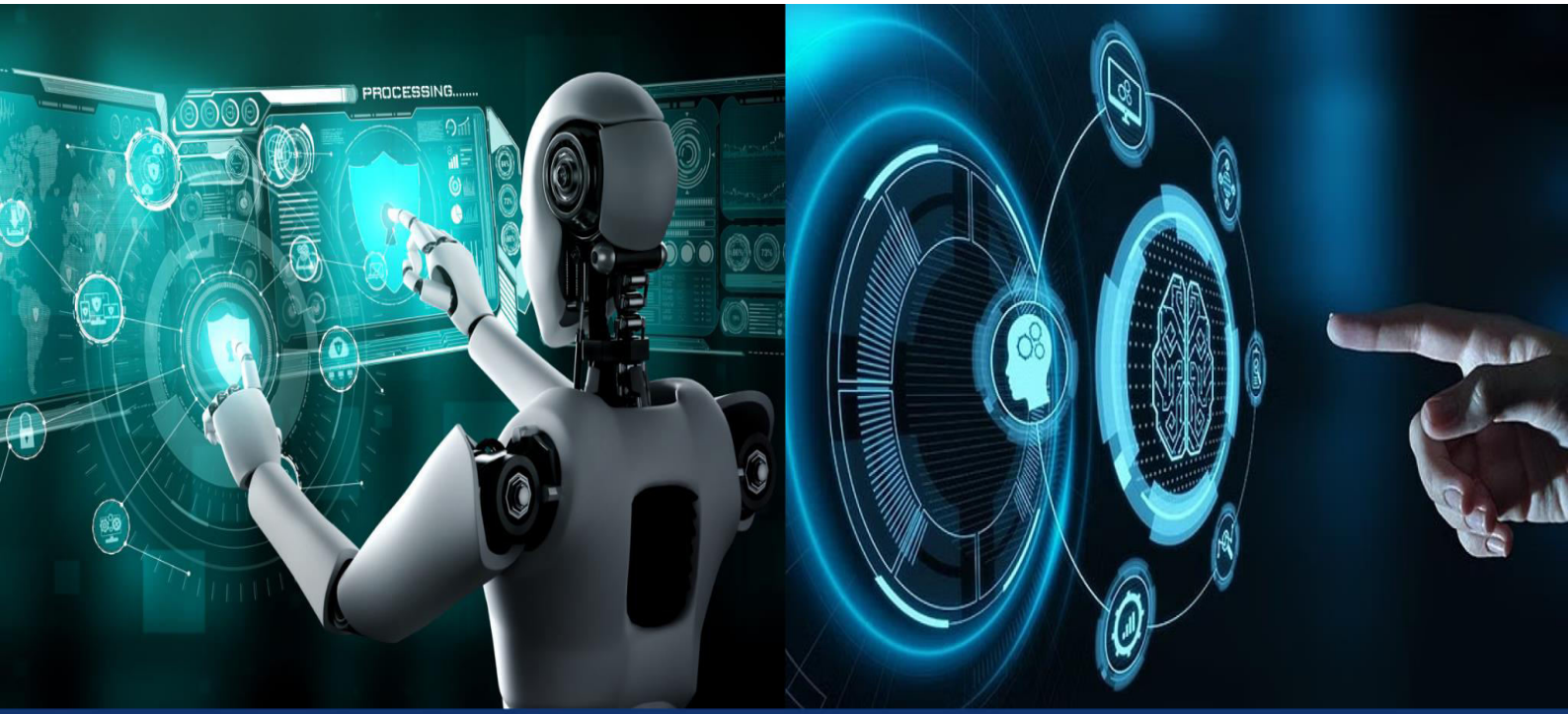
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771 **Volume 13, Issue 5, May 2025**

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305218

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Implementation of AI-Enabled Real Time Speech-To- Sign Language Converter with Animated Avatar

Arul R, Mathivanan V, Dhinesh Kumar M, S. Shruthi

Department of Information Technology, PSV College of Engineering and Technology, Krishnagiri, Tamil Nadu, India

ABSTRACT: In today's world, effective communication is a key necessity, yet individuals with hearing and speech impairments often face significant barriers, especially during real-time verbal interactions. This project aims to bridge that gap by developing an AI-enabled real-time speech-to-sign language converter using an animated avatar. The system captures spoken language, converts it into text using speech recognition, translates it if necessary, and finally maps the processed text to corresponding Indian Sign Language (ISL) signs which are then visualized using a 3D animated avatar. This seamless process enables non-signers to communicate effortlessly with the deaf community in real-time.

The project integrates technologies like Natural Language Processing (NLP), Text-to-ISL mapping, and avatar animation to produce expressive and contextually accurate sign language outputs. Support for both English and regional Tamil language input is incorporated, utilizing translation tools to maintain consistency in communication. The proposed solution is lightweight, responsive, and can be deployed in public environments like hospitals, railway stations, and schools, ensuring inclusivity and accessibility.

This project thus stands as a step toward digital empowerment and social equity, making verbal communication universally understandable through the power of artificial intelligence.

KEYWORDS: Artificial Intelligence, Speech Recognition, Indian Sign Language (ISL), Text-to-Sign Conversion, Real-Time Translation, Animated Avatar, Accessibility, Tamil-to-English Translation, NLP, Inclusive Communication.

I. INTRODUCTION

Communication is the foundation of human interaction. However, for the hearing and speech-impaired community, interacting with the world—especially in real-time spoken environments—remains a daily challenge. While sign language serves as their primary mode of communication, most people are unfamiliar with it, creating a communication gap. The emergence of Artificial Intelligence (AI), Natural Language Processing (NLP), and real-time translation technologies opens up new possibilities to solve this issue.

This project, titled **"AI-Enabled Real-Time Speech-to-Sign Language Converter with Animated Avatar"**, aims to create an inclusive system that captures speech, processes it intelligently, and converts it into sign language using a visually animated avatar. The goal is to enable seamless communication between hearing individuals and the hearing-impaired, particularly in real-time public or private interactions.

II. OBJECTIVE

- To design and implement a system that captures real-time speech and converts it into Indian Sign Language using a rule-based mapping approach.
- To integrate natural language processing techniques for accurate speech-to-text conversion, language

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305218

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

translation, and sentence restructuring suited for sign language output.

- To visualize the final output through an animated avatar that performs sign language gestures, enabling effective communication for hearing and speech-impaired individuals.

III. EXISTING AND PROPOSED SYSTEM

EXISTING SYSTEM

- Most existing systems use manual text input to generate sign language.
- Sign outputs are mostly static images or short video clips, not real-time or animated.
- Some tools support only American Sign Language (ASL), not Indian Sign Language (ISL).
- Real-time voice input processing is rarely implemented.
- Dependence on human interpreters for communication in many public spaces.

PROPOSED SYSTEM

- Converts real-time speech to sign language using AI technologies.
- Supports both English and Tamil voice inputs.
- Translates Tamil to English using a language translation module.
- Processes text to match ISL grammar and structure.
- Uses a 3D animated avatar to perform the signs for better clarity.
- Designed for use in public places like hospitals, stations, and school.

IV. ADVANTAGES

1. Real-time, accurate conversion of speech to ISL signs
2. Dual language input (English and Tamil) enhances usability.
3. No need for human interpreters, reducing dependency.
4. Can be scaled and deployed in real-world environments for inclusive communication .

SYSTEM FUNCTION

ARCHITECTURE DESIGN

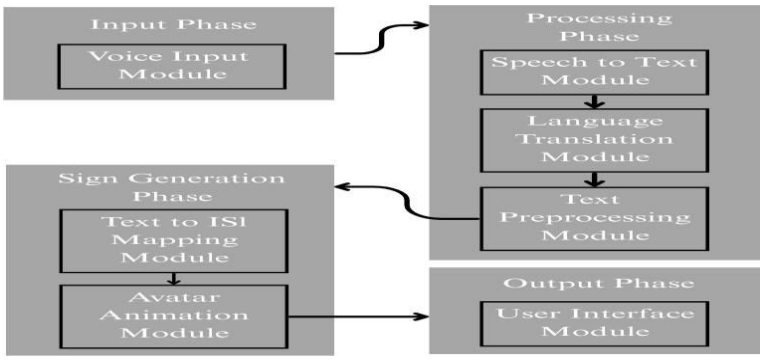


Figure 4.1 Proposed for architecture design

SYSTEM MODULES

They are six modules in this system.

- Speech-to-Text Conversion Module
- Language Translation Module

- Text Preprocessing Module
- Text-to-ISL Mapping Module
- Avatar Animation Module
- User Interface Module

Speech-to-Text Conversion Module:
The first module is the Speech-to-Text Conversion Module, which captures real-time speech input through a microphone and converts it into text using a speech recognition engine such as Google Speech API. It supports both English and Tamil speech inputs, enabling flexibility in usage.

Language Translation Module:
The second module is the Language Translation Module, which activates when the speech input is in Tamil. This module translates the Tamil text into English using a machine translation service like Google Translate, ensuring that the output is grammatically accurate and ready for further processing.

Text Preprocessing Module:
Next is the Text Preprocessing Module, which prepares the English text for ISL translation. This involves removing articles (a, an, the), auxiliary verbs, and restructuring the sentence to follow ISL grammar. It simplifies the sentence so that each word corresponds directly to a sign language gesture.

User Interface Module:
The User Interface Module provides a front-end platform for users to interact with the system. It allows users to select the input language, view the translated text, and watch the avatar's sign output. The interface is simple, intuitive, and accessible to users with minimal training.

Voice Input Module
The process begins with the **Voice Input Module**, which captures speech from the user using either a built-in or external microphone. This module is designed for real-time operation and includes mechanisms to reduce background noise, normalize volume, and ensure a clean audio stream. The moment the user speaks, the audio is buffered and forwarded to the next module for processing. This phase is critical, as the accuracy of downstream processing heavily depends on the clarity of the audio input.

Speech-to-Text Conversion Module
Once the audio is captured, it is sent to the **Speech-to-Text Conversion Module**. This module utilizes a speech recognition engine such as Google Speech API to convert spoken language into digital text. The system is capable of detecting and processing speech in both English and Tamil, making it suitable for users from diverse linguistic backgrounds. The speech recognition engine handles variations in pronunciation, tone, and accent, ensuring that the output text is as accurate as possible. If the speech is in English, the text proceeds to the next stage. If the speech is in Tamil, it is routed through the translation module before further processing.

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305218

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Language Translation Module

The **Language Translation Module** becomes active when the input speech is in Tamil. This module makes use of machine translation services such as the Google Translate API to convert Tamil text into English. Since the ISL mapping and animation modules rely on English input for processing, this translation step ensures consistency throughout the pipeline. The module is designed to maintain sentence meaning and structure while converting regional input to standard English, making the rest of the processing smoother and more reliable.

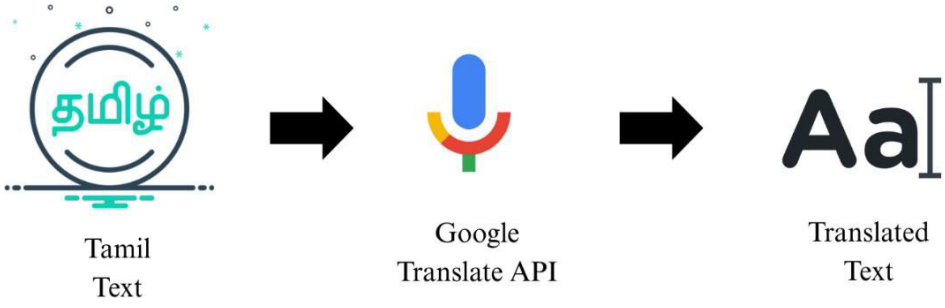


Figure Language Translation Module

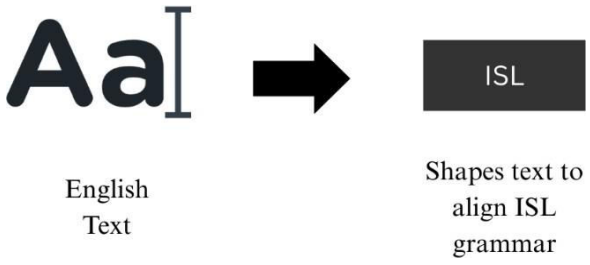


Figure Text Preprocessing Module

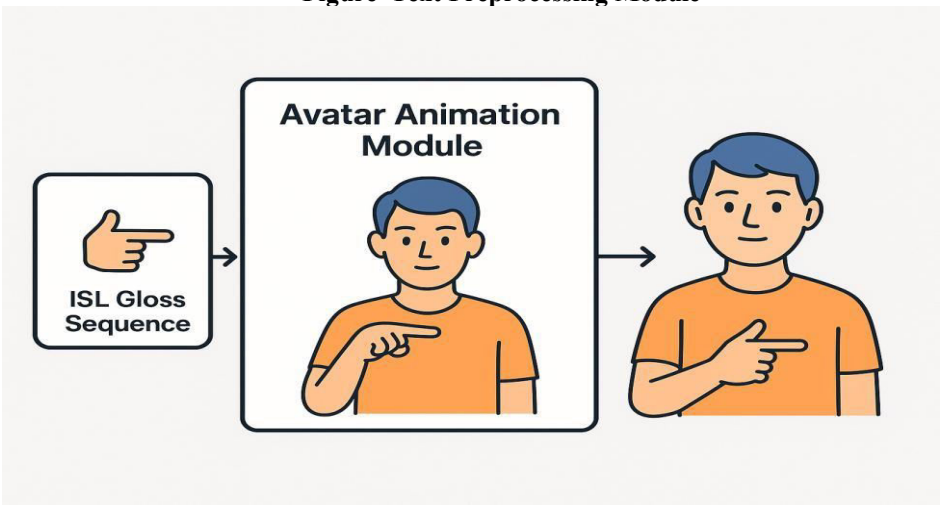


Figure Avatar Animation Module

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305218

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

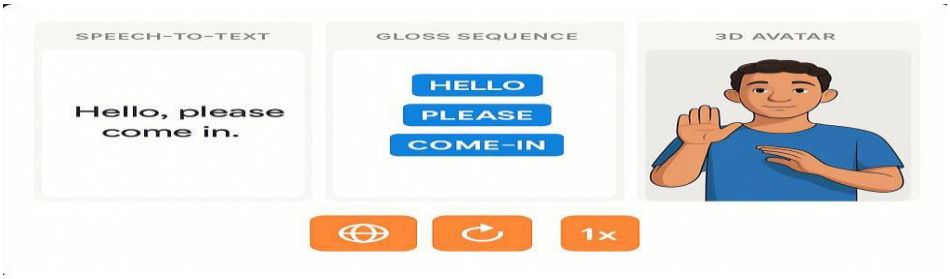


Figure User Interface Module

Sample Input-Output Test Case

To validate the system, a **Sample Input-Output Test Case** can be used. For instance, if the user says, "What is your name?", the system processes it to "NAME WHAT", maps it to the appropriate signs, and finally renders it through the avatar. This final output confirms that the system performs as intended, handling real-time voice input and producing a grammatically and visually correct ISL translation.

The avatar performs the signs sequentially with smooth transitions, making the meaning clear and accessible to a hearing-impaired user. The speech-to-text conversion is accurate and fast, and the preprocessing module effectively removes unnecessary words, preserving only the essential ISL-friendly content.

Figure Sample Input-Output Test Case

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305218

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

SPEECH PROCESSING

Speech processing refers to the analysis and manipulation of speech signals to extract meaningful information. In this system, the core objective of speech processing is to capture real-time human speech and convert it accurately into readable text, using automatic speech recognition (ASR) techniques. The system is designed to recognize both English and Tamil speech, ensuring inclusive communication for users across linguistic backgrounds.

The speech input is captured using a microphone connected to the system, and it is processed through APIs that specialize in converting speech to text. The processed text is then used in subsequent stages like translation, restructuring, and sign mapping. Any error or noise in this stage can cascade down and reduce the accuracy of the final output, hence speech processing must be highly reliable and efficient.

USER INTERFACE DESIGN

The user interface (UI) of the AI-enabled speech-to-sign language converter plays a crucial role in making the system intuitive, accessible, and efficient for end- users. Since the primary users of the system may include individuals with hearing impairments or their communication partners, the UI must be visually clear, responsive, and easy to navigate. The frontend design is developed using **Tailwind CSS**, a utility-first CSS framework known for its customizability and mobile-first responsive features.

DATABASE

The proposed system is designed to operate in real-time without requiring persistent storage for its core functionality. However, to improve modularity, scalability, and future expandability, a database module can be integrated to store gloss mappings, animation references, translation logs, and system metadata.

The central component of the database is the **Gloss Mapping Table**, which links each English word or phrase to its corresponding **ISL gloss**. In addition to the gloss, the database maintains references to **SiGML (Signing Gesture Markup Language)** files, which contain XML-based instructions for rendering the sign animations using a 3D avatar. These SiGML files describe each gesture's handshape, location, orientation, and movement, making the animation process dynamic and scalable.

Furthermore, the system incorporates the **Stanford NLP Parser (.jar)** to analyze the grammatical structure of the input sentence during the preprocessing stage. The parser helps identify parts of speech and sentence hierarchy, enabling the system to restructure complex sentences into a simplified format suitable for ISL conversion. This ensures that the gloss mapping is applied to a properly interpreted and contextually accurate sentence, improving the overall reliability of the system. The database also supports optional components like:

- **Translation Logs**: storing past voice-to-sign conversions for debugging or audit purposes.
- **User Profiles**: for saving language preferences and accessibility settings.
- **Feedback Data**: to gather user input and improve system accuracy over time.

Refer Table 5.3: Gloss-Action Mapping Table, which includes fields like Word, ISL Gloss, SiGML Reference, and gesture metadata for smooth animation handling.

Word	ISL Gloss	SiGML File	Handshape	Location	Movement
Hello	HELLO	hello.sigml	Flat Hand	Chest	Forward Wave
Help	HELP	help.sigml	Flat + Fist	Torso	Lift Motion
Where	WHERE	where.sigml	Pointing Index	Front	Side-to-Side

Table Gloss-Action Mapping Table

Sign Animation Metadata using SiGML

In the proposed system, each Indian Sign Language (ISL) gloss is linked to a corresponding **SiGML**

(Signing Gesture Markup Language) file. SiGML is an XML-based markup language used to define and render sign language gestures in avatar animation engines. It allows precise control over the handshape, location, movement, palm orientation, and non-manual expressions like facial gestures.
Each gloss in the database is associated with a unique SiGML file that contains the gesture instructions for the 3D avatar. These files are interpreted by the animation engine to produce realistic and fluid signs.
The **Gloss-Action Mapping Table** in the database includes a column referencing the file path or ID of the associated SiGML file. This structured format allows for scalable animation and easy updates without modifying the animation codebase.

V. RESULT

The result of this project is a user-friendly application leveraging Artificial Intelligence to detect counterfeit products, thereby empowering non-tech-savvy customers to identify fake items quickly and accurately. By comparing input images of received products with a database of original product images using the SIFT algorithm, the application determines whether a product is genuine or counterfeit. This automatic image recognition and classification system not only enhances consumer confidence in online purchases but also contributes to combating product piracy by enabling users to make informed decisions and avoid potentially dangerous counterfeit products, particularly in the case of medical items where safety is paramount. Through the utilization of AI-driven technology, this project aims to address the pervasive issue of counterfeit products in the online retail industry, ensuring safer and more secure shopping experiences for consumers worldwide.

VI. CONCLUSION

This project represents a significant advancement in combating the proliferation of counterfeit products, particularly in the online retail industry. By harnessing the power of Artificial Intelligence, specifically through image recognition and object detection using the SIFT algorithm, this solution empowers non-tech-savvy consumers to identify fake products with ease and precision. By leveraging a database of original product images, the system compares input images to detect discrepancies and classify products as either genuine or counterfeit. This approach not only enhances consumer safety and confidence but also contributes to the broader effort of combating product piracy and protecting consumers from potentially harmful counterfeit goods, especially in critical sectors like medicine. With its user-friendly interface and robust detection capabilities, this project serves as a valuable tool in the ongoing fight against counterfeit products, ultimately safeguarding consumer interests and promoting trust in online retail.

REFERENCES

1. Li, Zhe, et al. “Fast Logo Detection and Recognition in Document Images.” 2020 20th International Conference on Pattern Recognition, 2020.
2. Shull, F., Boehm, V.B., Brown, A., Costa, P., Lindvall, M.,Port, D., Rus, I., Tesoriero, R., and Zelkowitz, M. 2021.What We Have Learned About Fighting Defects.In Proceedings of the Eighth International Software Metrics Symposium, 249-258, 2021.
3. Dreiseitl, Stephan, and Lucila Ohno-Machado. “Logistic Regression and Artificial Neural Network Classification Models: a Methodology Review”, 2022.
4. Pathak, Abhishek, et al. “Implicit and Explicit Identification of Counterfeit Brand Logos Based on Logotype Transposition.” Journal of Product & Brand Management, 2019.
5. Eka Dyar Wahyuni and Arif Djunaidy, “Fake Review Detection from a Product Review Using Modified Method of Iterative Computation Product”, January 2019 Research Gate.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details