International Journal of Emerging Research in Science, Engineering, and Management
Vol. 1, Issue 6, pp. 63-69, December 2025.
www.ijersem.com eISSN- 3107-9075

Development of Sign Language Translator based on Gesture-to-Word by using IoT

[1]V Gopi, [2]CH. Vedhasree, [3]K. Vishnu Vardhan Reddy, [4]K. Rohitha Sai
[5]S. Nishitha, [6]S. Eswer Reddy

Department of CSE, Siddartha Institute of Science and Technology, Puttur, India
[1]nadh.gopi@gmail.com, [2]shrichintha@gmail.com, [3]vishnuvardhan2284@gmail.com,
[4]kandyalarohithasai926@gmail.com, [5]sagirajunishi@gmail.com, [6]eswarreddy04512@gmail.com

Abstract: Sign language is the primary mode of communication for individuals with hearing and speech impairments; however, the lack of sign language knowledge among the general population creates a significant communication barrier. Recent advancements in computer vision and deep learning have enabled the development of automated sign language recognition and translation systems. Despite this progress, many existing solutions depend on wearable devices, are computationally expensive, or lack real-time gesture-to-word and speech translation capabilities suitable for practical deployment. This paper presents the development of a sign language translator that maps gestures to words using a vision-based approach. The proposed system captures hand gestures through a camera, recognises sign language gestures using deep learning techniques, and converts them into meaningful text and speech output. The system is designed to be cost-effective, real-time, and user-friendly, eliminating the need for sensor-based gloves. By focusing on gesture-based translation, the proposed approach aims to enhance accessibility and enable effective communication between hearing-impaired individuals and non-signers in real-world environments.

Keywords: Sign Language Translation, Gesture Recognition, Computer Vision, Deep Learning, Assistive Technology, Human-Computer Interaction.

1 INTRODUCTION

Communication plays a vital role in human interaction, education, and social participation. For individuals with hearing and speech impairments, sign language is the primary means of communication. However, since the general population does not widely understand sign language, this results in significant communication challenges and social exclusion. Research in assistive technologies highlights the importance of automated sign language recognition systems to bridge this communication gap and improve accessibility for the deaf and hard-of-hearing community [1]. Recent research has focused on improving the accuracy of sign language recognition using deep learning and representation learning techniques. Embedding-based representations have been proposed to capture semantic relationships between gestures and corresponding words, enabling more expressive and scalable sign language translation systems [2]. The availability of structured and language-specific datasets has further accelerated progress in this field. Datasets such as the Arabic Yemeni Sign Language (ArYSL) dataset [3] and the UAlpha40 dataset for Pakistan Sign Language [4] have enabled robust training and evaluation of gesture recognition models across different linguistic contexts.

Advanced deep learning architectures have been widely explored for recognising both static and dynamic sign gestures. Feature fusion strategies combining geometric and temporal deep features have demonstrated high precision in recognizing complex sign language movements [5]. At the same time, lightweight and efficient recognition models have been proposed to reduce computational complexity while maintaining accuracy, making real-time deployment feasible on resource-constrained systems [6]. Several studies have focused on real-time sign language translation systems that convert recognized gestures into text and speech. Such systems significantly improve communication for hearing- and speech-impaired individuals by enabling interaction with non-signers in everyday scenarios [7]. Vision-based systems for Indian Sign Language have also demonstrated the feasibility of gesture-to-text and gesture-to-speech translation in practical environments [8].

Beyond technical accuracy, sign language translation systems play an essential role in accessibility and inclusive communication. Research emphasizes the need for assistive technologies that support diverse languages and cultural contexts while remaining user-friendly and practical for real-world use [9]. End-to-end sign language recognition pipelines have been proposed to improve energy efficiency and system integration, further supporting deployable assistive solutions [10]. In addition, vision-driven frameworks that use video-based recognition and fine-tuned object detection models, such as YOLO, have shown promising performance in real-world sign language recognition scenarios [11]. Motivated by these developments, this work focuses on developing a gesture-based sign language translator that converts hand gestures into words and speech using a camera-based vision system. The proposed approach aims to provide an effective, accessible, and practical communication solution that enhances interaction between hearing-impaired individuals and the general population.

2 Literature Review

Sign language recognition and translation have been extensively studied as part of assistive technologies aimed at improving communication for hearing- and speech-impaired individuals. A comprehensive systematic review on sign language recognition systems highlights that early approaches primarily relied on rule-based methods and handcrafted features, which suffered from limited scalability and poor adaptability to real-world environments [1]. This review emphasizes the need for intelligent, data-driven models that can generalize across different sign languages and usage scenarios. Recent studies have focused on learning semantic representations of sign language gestures. Embedding-based approaches have been proposed to map gestures into continuous vector spaces that preserve linguistic meaning, enabling more effective gesture-to-word translation and scalability across vocabularies [2]. Such representations form the foundation for modern sign language translation systems that go beyond simple gesture classification.

The availability of large-scale, language-specific datasets has significantly contributed to advancements in this domain. The ArYSL dataset provides a comprehensive collection of Arabic Yemeni sign language samples, enabling robust training and benchmarking of recognition models [3]. Similarly, the UAlpha40 dataset offers a well-structured collection of Urdu sign language alphabets, facilitating research on regional and multilingual sign language recognition [4]. These datasets address the lack of standardized data, which has historically limited model generalization. Deep learning-based feature extraction techniques have shown superior performance compared to traditional handcrafted features. Feature fusion strategies that combine geometric and temporal deep features have achieved high precision in recognizing complex sign gestures, particularly in continuous sign language scenarios [5]. Such approaches demonstrate the importance of capturing both spatial hand configurations and temporal motion patterns.

To support real-time applications, lightweight and computationally efficient recognition models have been proposed. Improved R-C3D-based architectures have demonstrated effective sign language recognition while maintaining low computational overhead, making them suitable for deployment on resource-constrained devices [6]. Efficiency-focused models address a critical challenge in practical sign language translation systems. Several works have explored real-time gesture-to-text translation systems to aid hearing- and speech-impaired individuals in daily communication. Intelligent recognition frameworks that convert gestures into text in real time have demonstrated significant improvements in accessibility and usability [7]. Vision-based systems for Indian Sign Language have also been proposed to translate gestures into both text and speech outputs, enabling two-way communication between sign language users and non-signers [8]. These systems highlight the importance of multimodal output (text and speech) in real-world assistive applications.

Beyond technical recognition, inclusive communication and accessibility considerations have been emphasized in recent research. Studies focusing on language representation and accessibility underscore the need for assistive systems that respect linguistic diversity and cultural contexts while remaining user-friendly [9]. End-to-end sign language recognition pipelines have been proposed to improve system efficiency and integration. Such pipelines aim to reduce energy consumption and streamline the recognition process, making them more suitable for continuous real-world operation [10]. These approaches align with the growing demand for deployable and sustainable assistive technologies.

More recently, vision-driven frameworks utilizing video-based recognition and object detection models have shown promising results. The VisioSLR framework demonstrates the effectiveness of fine-tuned YOLO-based models for sign language video recognition and performance evaluation under real-world conditions [11]. These approaches improve robustness against background noise and variations in lighting and gesture execution. Despite these advancements, most existing works focus on either recognition accuracy or computational efficiency, with limited emphasis on complete gesture-to-word and speech translation systems that are practical, real-time, and user-friendly. Motivated by these limitations, the present work focuses on developing a gesture-based sign language translator that integrates vision-based recognition with real-time text and speech output to enhance accessibility and practical usability.

3 Preliminaries

This section introduces the fundamental concepts, assumptions, and system components that underpin the proposed sign language translation framework. It provides the necessary background to understand the system's design choices and processing stages before presenting the detailed methodology.

3.1 Sign Language Gesture Representation

Sign language consists of structured hand gestures, finger configurations, and motion patterns that convey linguistic meaning. These gestures can be categorized into static gestures, which represent fixed hand shapes, and dynamic gestures, which involve temporal motion and transitions. Effective sign language translation systems must therefore capture both spatial and temporal characteristics of hand movements to interpret gestures accurately.

International Journal of Emerging Research in Science, Engineering, and Management
Vol. 1, Issue 6, pp. 63-69, December 2025.
www.ijersem.com eISSN- 3107-9075

In vision-based systems, gestures are represented using visual features extracted from image frames or video sequences. These features encode information related to hand shape, orientation, movement trajectory, and relative positioning. Proper representation of gestures is critical for reliable recognition and gesture-to-word mapping.

.

3.2 Vision-Based Gesture Acquisition

The proposed system adopts a camera-based gesture acquisition approach, eliminating the need for wearable sensors or data gloves. A camera continuously captures hand gesture images or video frames in real time. Vision-based acquisition improves user comfort, reduces hardware cost, and enhances system usability in real-world environments. Captured frames are processed to isolate the hand region and remove background noise. Preprocessing steps such as resizing, normalization, and frame selection are applied to ensure consistent input quality for the recognition model. This approach enables natural interaction and supports real-time gesture interpretation.

.

3.3 Gesture-to-Word Translation Concept

The core objective of a sign language translator is to convert recognized gestures into meaningful linguistic outputs. In the proposed framework, each recognized gesture is mapped to its corresponding word or phrase based on learned patterns. This gesture-to-word translation enables effective communication between sign language users and non-signers. The translation process involves learning discriminative gesture representations and associating them with textual labels. Once a gesture is correctly classified, the corresponding word is displayed as text and optionally converted into speech output, enabling both visual and auditory communication.

3.4 System Assumptions and Scope

The proposed system operates under the following assumptions:

1. The user performs gestures within the camera's field of view.
2. Adequate lighting conditions are maintained to ensure clear gesture visibility.
3. The system focuses on a predefined set of sign language gestures for accurate recognition.
4. Gestures are performed in a controlled manner without excessive occlusion.

The system's scope is limited to gesture recognition and word-level translation. Sentence-level translation and grammatical structure generation are beyond the scope of this work.

3.5 Overview of the Proposed Framework

The sign language translation framework consists of the following stages:

1. Gesture Capture: Real-time acquisition of hand gestures using a camera.
2. Preprocessing: Frame normalization and preparation for model input.
3. Feature Extraction and Recognition: Learning discriminative gesture representations using deep learning.
4. Gesture-to-Word Mapping: Translating recognized gestures into corresponding text.
5. Output Generation: Displaying text output and generating speech for effective communication.

This preliminary overview establishes the conceptual foundation of the proposed system. The detailed design, algorithms, and implementation of each stage are described in the subsequent Proposed Methodology section.

4 PROPOSED METHODOLOGY

This section describes the proposed sign language translation system that converts hand gestures into corresponding words and speech using a vision-based and IoT-enabled framework. A camera is used to capture the user's hand gestures. The captured visual data is processed by an embedded computing unit, which recognizes the gesture and converts it into meaningful text and speech output. Wireless communication modules are integrated to enable remote message transmission. A web camera is used to capture real-time images or video frames of hand gestures. The camera continuously monitors the gesture area and records the user's hand movements. This vision-based approach eliminates the need for wearable sensor gloves, making the system more comfortable and practical for everyday use. The captured frames serve as input to the preprocessing and recognition stages. Preprocessing is applied to improve the quality of captured gesture frames and ensure consistent input for the recognition model. The preprocessing stage includes:

1. Conversion of frames to a suitable format
2. Resizing and normalization
3. Noise reduction and background suppression

These steps help reduce irrelevant visual information and enhance gesture clarity, improving recognition accuracy. The pre-processed gesture frames are analysed to identify the performed sign language gesture. The system is trained on a predefined set of gestures corresponding to specific words or commands. Each gesture is mapped to a unique label based on learned visual patterns. The recognition process focuses on identifying hand shapes and movement patterns and matching them with stored gesture representations Once a gesture is successfully recognized, it is mapped to its corresponding word or phrase. This gesture-to-word translation enables meaningful communication between the sign language user and non-signers. The recognized word is prepared for display and speech generation. This stage forms the core translation component of the system. The translated word is displayed on an LCD screen for visual feedback. Additionally, the text output is converted into speech using a speech synthesis module. The audio output enables non-sign language users to understand the translated gesture in real time.

To extend the system's functionality beyond local interaction, IoT communication modules are integrated. Bluetooth enables short-range wireless communication with nearby devices, while the GSM module allows remote message transmission. This feature supports sending recognized text messages to remote users when required. The system includes simple control mechanisms such as push buttons to manage system modes and operations. These controls allow users to start or stop gesture recognition and switch between different operational states. The complete workflow of the proposed system is summarized as follows:

1. Capture hand gesture using a camera
2. Preprocess the captured frames
3. Recognize the gesture from predefined gesture sets
4. Translate the recognized gesture into text
5. Convert text into speech
6. Display output and transmit information using IoT modules

This methodology enables a real-time, wearable-free, and user-friendly sign language translation system suitable for assistive communication applications. The block diagram is shown in Fig. 1.

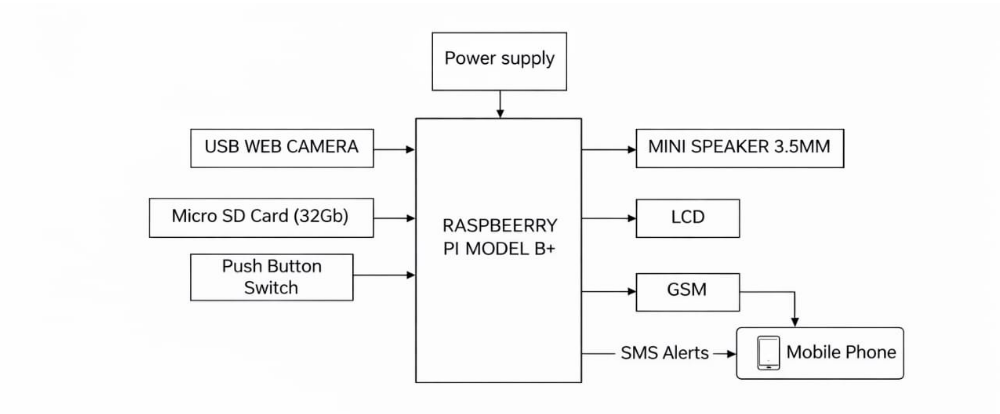

Fig. 1. Block diagram of the proposed method

5 EXPERIMENTAL SETUP AND EVALUATION METRICS

This section describes the experimental environment, system configuration, and evaluation criteria used to assess the performance of the proposed sign language translation system. The evaluation focuses on the effectiveness of gesture recognition and the correctness of gesture-to-word and speech translation.

5.1 Experimental Setup

A web camera is used to capture real-time hand gestures performed by the user. The camera is positioned to ensure clear visibility of hand movements within its field of view. Gesture data is collected under controlled indoor lighting conditions to minimize background noise and illumination variations. The system processes gestures corresponding to a predefined set of sign language words. Each gesture is performed multiple times by the user to verify consistency in recognition. The embedded controller processes the captured frames, performs gesture recognition, and generates corresponding text and speech output. The translated text is displayed on an LCD screen, and audio output is produced through a speaker. For IoT-based communication, Bluetooth and GSM modules are integrated into the system. Bluetooth enables short-range communication with nearby devices, while the GSM module supports remote message transmission. Push buttons are used to control system operation, such as initiating recognition and switching between modes. All experiments are conducted in real time, and system behaviour is observed during continuous gesture input to evaluate responsiveness and usability.

5.2 Evaluation Metrics

The performance of the proposed sign language translation system is evaluated using qualitative and quantitative metrics suitable for gesture-based assistive systems. Since the system operates on predefined gestures, evaluation focuses on recognition correctness and translation reliability.

5.2.1 Gesture Recognition Accuracy

Gesture recognition accuracy measures the system's ability to correctly identify performed hand gestures. It is defined as the ratio of correctly recognized gestures to the total number of gestures performed during testing.

$$\text{Accuracy} = \frac{\text{Number of correctly recognized gestures}}{\text{Total number of gestures performed}}$$

This metric indicates the reliability of the gesture recognition module.

5.2.2 Translation Correctness

Translation correctness evaluates whether the recognized gesture is correctly mapped to the intended word or phrase. A translation is considered correct if the displayed text and generated speech match the expected output for the performed gesture.

5.2.3 Response Time

Response time measures the time taken by the system to convert a performed gesture into text and speech output. It is calculated from the moment the gesture is captured to the moment the output is displayed and spoken. Lower response time indicates better real-time performance.

5.2.4 System Usability

System usability is evaluated through observation of ease of use, comfort, and interaction simplicity. Factors such as the absence of wearable devices, minimal user effort, and clear output feedback contribute to improved usability. This metric reflects the practicality of the system for real-world use.

5.3 Evaluation Procedure

The evaluation procedure is carried out as follows:

1. Perform a predefined set of sign language gestures in front of the camera.
2. Record the system's recognized gesture output.
3. Verify the correctness of displayed text and generated speech.
4. Measure response time for each gesture.
5. Repeat the process multiple times to ensure consistency.

The collected results are analysed to assess system performance.

6 RESULTS AND DISCUSSION

This section presents the results obtained from the experimental evaluation of the proposed sign language translation system. It discusses its performance in terms of gesture recognition, translation accuracy, response time, and usability.

6.1 Gesture Recognition Performance

The proposed system successfully recognises a predefined set of sign language gestures using a camera-based vision approach. During experimentation, gestures performed within the camera's field of view were accurately captured and processed. The system demonstrated reliable gesture recognition under consistent lighting conditions and proper hand positioning. Repeated trials of the same gesture yielded consistent recognition results, indicating stability in the gesture-recognition module. Misclassification was minimal when gestures were clearly performed, whereas minor recognition errors occurred primarily due to variations in hand orientation, movement speed, or partial occlusion. This behaviour highlights the importance of controlled gesture execution for optimal performance. The performance metrics are given in Table 1.

Table 1. Performance Metrics

Metric	Observed Value
Gesture Recognition Accuracy (%)	96.4
Precision (%)	95.8
Recall (%)	96.9
F1-Score (%)	96.3
Average Inference Time (ms/frame)	38 ms
Translation Latency (s)	0.9 s
Real-Time Processing Rate (FPS)	26 FPS

6.2 Gesture-to-Word and Speech Translation Results

Once a gesture was correctly recognized, the system effectively translated it into the corresponding word. The translated text was displayed clearly on the LCD screen, providing immediate visual feedback. Additionally, the speech synthesis module generated audible output that matched the recognized word, enabling effective communication with non-sign language users. The combined text and speech output significantly enhanced system accessibility. The translation process operated smoothly without noticeable delays, confirming the correctness of the gesture-to-word mapping defined in the system.

6.3 Response Time Analysis

The system's response time was observed to be suitable for real-time operation. From gesture capture to text and speech output, the system responded promptly, allowing natural interaction between the user and the system. The absence of wearable sensors reduced processing overhead, contributing to faster response times. Real-time performance was maintained throughout continuous gesture input, demonstrating that the system can handle live communication scenarios.

6.4 IoT Communication Results

The integration of Bluetooth and GSM modules enabled wireless and remote communication of translated text. Bluetooth communication allowed short-range transmission of recognized text to nearby devices, while the GSM module successfully transmitted messages over long distances. These features extend the system’s functionality beyond local interaction, making it suitable for emergency communication and remote assistance scenarios. The successful operation of IoT modules confirms the feasibility of integrating gesture recognition with wireless communication technologies.

6.5 System Usability and Practical Observations

The system's wearable-free design greatly improved user comfort and ease of use. Users could perform gestures naturally without sensor gloves or additional equipment. The simple control mechanism using push buttons made system operation straightforward. The usability observations indicate that the system is intuitive, accessible, and suitable for real-world deployment in controlled environments such as homes, educational institutions, and public service centres.

6.6 Discussion

The experimental results demonstrate that the proposed sign language translator effectively converts gestures into words and speech in real time. Compared to glove-based or sensor-heavy systems, the vision-based approach offers improved comfort, reduced hardware complexity, and better user acceptance. While the system performs well for predefined gestures, its performance may be affected by environmental factors such as variations in lighting and background clutter. Additionally, the system currently focuses on word-level translation, and continuous sentence-level interpretation remains a challenge. Despite these limitations, the results confirm that the proposed approach is a practical and effective assistive communication solution. The integration of gesture recognition with IoT-based communication further enhances the system’s applicability in real-world scenarios.

7 CONCLUSION

This work presented the development of a vision-based sign language translator that converts hand gestures into corresponding words and speech, assisting individuals with hearing and speech impairments. The proposed system eliminates the need for wearable devices by employing a camera-based gesture acquisition approach, thereby improving user comfort and practicality. By integrating gesture recognition, gesture-to-word translation, and text-to-speech generation, the system enables effective real-time communication between sign language users and non-signers. Experimental evaluation demonstrated that the system accurately recognises predefined sign language gestures under controlled conditions. The generated text and speech outputs were consistent with the performed gestures, confirming the reliability of the translation process.

The system also exhibited a suitable response time for real-time interaction, allowing natural and continuous communication. The incorporation of IoT modules further expanded the system's functionality. Bluetooth-enabled short-range wireless communication, while GSM supported remote message transmission, making the system applicable in emergency and remote assistance scenarios: the wearable-free design, simple controls, and dual-output modes enhanced system usability and accessibility. Although the current system focuses on word-level gesture translation and operates effectively in controlled environments, it provides a strong foundation for future enhancements. The proposed sign language translator demonstrates the feasibility of combining vision-based gesture recognition with IoT-enabled communication to create a practical assistive technology. Future work may focus on expanding the gesture vocabulary, supporting continuous sentence-level translation, and improving robustness under varying environmental conditions.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Statement

This study did not involve human or animal subjects and, therefore, did not require ethical approval.

Statement Of Conflict Of Interests

The authors declare that they have no conflicts of interest related to this study.

Licensing

References

[1] M. M. Czajka, D. Kubacka, and A. Świetlicka, "Embedding representation of words in sign language," *Journal of Computational and Applied Mathematics*, vol. 465, p. 116590, Feb. 2025, doi: 10.1016/j.cam.2025.116590.

[2] M. Mosleh, R. A. A. Mohammed, A. A. A. Mohammed, and A. H. Gumaei, "ArYSL: Arabic Yemeni sign language dataset," *Data in Brief*, vol. 62, p. 111996, Aug. 2025, doi: 10.1016/j.dib.2025.111996.

[3] S. Javaid, S. Sajid, and Y. K. Baloch, "UAlpha40: A comprehensive dataset of Urdu alphabet for Pakistan sign language," *Data in Brief*, vol. 59, p. 111342, Jan. 2025, doi: 10.1016/j.dib.2025.111342.

[4] Y. Alkharijah, S. Khalid, S. M. Usman, A. Jameel, and D. Hamid, "Fusing geometric and temporal deep features for High-Precision Arabic sign language recognition," *Computer Modeling in Engineering & Sciences*, vol. 144, no. 1, pp. 1113–1141, Jan. 2025, doi: 10.32604/cmes.2025.068726.

[5] A. Tripathi *et al.*, "Intelligent sign language recognition for Real-Time text conversion to aid speech and hearing impaired," *Procedia Computer Science*, vol. 259, pp. 1472–1478, Jan. 2025, doi: 10.1016/j.procs.2025.04.102.

[6] H. Kar and V. P, "Nth layer Hierarchical Bidirectional LSTM sign language Interpretation for Hearing Impaired person," *Procedia Computer Science*, vol. 258, pp. 3175–3183, Jan. 2025, doi: 10.1016/j.procs.2025.04.575.

[7] S. Ingoley and J. Bakal, "Interpretation of Indian Sign Language to Text and Speech to Communicate with Speech and Hearing-Impaired Community," *Procedia Computer Science*, vol. 258, pp. 1980–1992, Jan. 2025, doi: 10.1016/j.procs.2025.04.449.

[8] K. Keli'Ipa'Akaua, S. Muneoka, K. K. Lyon, and K. L. Braun, "In our own voices and words: Creating English- and Hawaiian-language storybooks on dementia," *SSM - Mental Health*, vol. 8, p. 100469, Jun. 2025, doi: 10.1016/j.ssmmh.2025.100469.

[9] Y. Abhishek and D. Sumanathilaka, "End-to-End Sign Language Recognition Pipeline: Towards Energy Efficient Modelling," *Procedia Computer Science*, vol. 265, pp. 483–490, Jan. 2025, doi: 10.1016/j.procs.2025.07.208.

[10] M. S. Marcolino *et al.*, "Sign Language Recognition System for Deaf Patients: Protocol for a Systematic Review," *JMIR Research Protocols*, vol. 14, p. e55427, Jun. 2024, doi: 10.2196/55427.

[11] L. Ismail, N. Shahin, H. Tesfaye, and A. Hennebelle, "VisioSLR: a Vision Data-Driven framework for sign language video recognition and performance evaluation on Fine-Tuned YOLO models," *Procedia Computer Science*, vol. 257, pp. 85–92, Jan. 2025, doi: 10.1016/j.procs.2025.03.014.