



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

An Intelligent Predictive framework for Real-Time Arabic Sign Language Translation through Multimodal Deep Learning and Cognitive Vision

Munay Said Ambani¹, Amira Muftah Abubaker Abushrida², Farhat Mahmoud Embarak³,

Esraa Ahmed Sanousi⁴, Amina Abdulrahman Mohammed⁵, Narmin salah AlKhalidi⁶

Faculty of Engineering, Ajdabiya University Ajdabiya, Libya¹

Faculty of Electrical & Electronic, Engineering, Bright Star University AlBurayqah, Al-Burayqah, Libya¹

Department of Information Technology, Faculty of Computing and Information Technology, Ajdabiya University, Ajdabiya, Libya²

Faculty of Computing and Information, Technology, Ajdabiya University Libya³

Faculty of Computing, Universiti, Teknologi Malaysia, Johor Bahru Malaysia³

<https://orcid.org/0000-0002-9225-9532>

Department of Information Technology, Faculty of Computing and Information Technology, Ajdabiya University Ajdabiya, Libya⁴

Department of Information Technology, Faculty of Computing and Information Technology, Ajdabiya University, Ajdabiya, Libya⁵

Department of Information Technology, Faculty of Computing and Information Technology, Ajdabiya University, Ajdabiya, Libya⁶

ABSTRACT: Sign language serves as the primary communication medium for millions of deaf and hard-of-hearing individuals worldwide, yet language barriers continue to limit their full social participation. Developing intelligent systems capable of accurately translating sign language into spoken or written text is therefore essential to achieving genuine accessibility and inclusion. This study presents an intelligent predictive framework for real-time Arabic Sign Language (IPF-ASL) translation, developed to enable seamless, natural communication between deaf and hearing individuals. The framework architecture combines state-of-the-art deep learning and computer vision technologies. A Vision Transformer module is employed to extract detailed spatial features from video frames, capturing subtle hand movements, finger orientations, and gesture dynamics. A Temporal Convolutional Network (TCN) processes sequential patterns to interpret temporal dependencies across consecutive signs, while a Predictive Module forecasts the upcoming gesture before completion, significantly reducing translation latency. These components are integrated with a MobileNetV2-based Convolutional Neural Network (CNN) optimized for lightweight, high-accuracy classification of Arabic sign gestures. The framework was trained using the RGB Arabic Sign Language Dataset, containing 24,800 labeled images representing 31 Arabic sign classes, including complex forms like La, Al, and Taa Marbuta. Comprehensive preprocessing—normalization, augmentation, and transfer-learning-based feature extraction—ensured balanced, high-quality inputs. Training was carried out in Python (TensorFlow/Keras) on Google Colab, with an 80/20 train-validation split for robust evaluation. Experimental results demonstrated a 97.8% classification accuracy on static Arabic signs and an average response latency below 150 milliseconds during real-time translation. The predictive module further improved user interaction by displaying text predictions 0.3 seconds before gesture completion, significantly enhancing conversational fluidity. User testing confirmed that the system is highly accurate, responsive, and intuitive, making it suitable for educational, healthcare, and public-service applications.

KEYWORDS: Arabic Sign Language, Computer Vision, Deep Learning Models, Improving Interaction for the Deaf and Hard of Hearing

I. INTRODUCTION

The rapid advancement of digital technologies and artificial intelligence (AI) has led to a fundamental transformation in human communication methods. In particular, computer vision and deep learning have enabled machines to interpret complex visual patterns, paving the way for intelligent assistive systems. Among these, sign language recognition (SLR) has emerged as a critical research domain aimed at improving communication for deaf and hard-of-hearing (DHH) individuals by translating visual gestures into understandable textual or spoken formats [1].

Despite technological progress, the DHH community continues to face significant challenges in social integration due to the linguistic gap between sign language users and the hearing population. Although sign language is the most natural and effective mode of communication for DHH individuals, everyday interactions are often limited by the lack of sign language proficiency among the general public. This limitation underscores the need for automated translation systems that can bridge this communication gap using intelligent and scalable solutions [2].

Recent advances in computer vision and deep learning have made it possible to design systems capable of analyzing hand gestures and body movements in real time. Vision-based gesture and sign recognition systems commonly employ convolutional neural networks (CNNs) for spatial feature extraction, along with recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and transformer architectures for modeling temporal dependencies in sign sequences. These approaches have demonstrated strong performance in recognizing complex hand gestures under varying conditions [2], [3].

Motivated by these developments, this project aims to develop an intelligent real-time application for translating Arabic Sign Language (ArSL) into written text and audible speech using computer vision and deep neural networks. Arabic Sign Language remains underrepresented in comparison with other sign languages, and existing studies emphasize the importance of language-specific datasets and models to achieve reliable recognition performance [4], [6],[8].

Furthermore, this project extends beyond conventional sign language translation by incorporating motion prediction mechanisms that anticipate the next sign before it is fully completed[7],[10]. By analyzing spatial-temporal patterns of hand movements, predictive models can reduce system latency and deliver a smoother and more natural communication experience. Such early prediction techniques are increasingly recognized as essential for real-time sign language recognition and interactive assistive systems [5][9].

By integrating visual feature extraction, temporal modeling, transformer-based architectures, and motion prediction techniques, the proposed system aims to bridge the communication gap between DHH individuals and the wider community. The project contributes to enhancing digital inclusion and accessibility, particularly in educational, medical, and administrative environments within Arabic-speaking regions such as Libya.

II. RELATED WORK

In recent years, a new research trend has emerged that focuses on predictive models in sign language processing, aiming to anticipate the upcoming movement before the sign is fully completed. This helps reduce response time and improve the realism of interaction. In a recent study, [11] presented a model for predicting Chinese Sign Language gestures based on neural networks specialized in analyzing sequential gestures. The study achieved an accuracy of 91.5% in real-time prediction of the next motion frame, reflecting the importance of integrating temporal analysis into translation systems. Likewise, [12] developed a model based on KFM-CNN to predict static and dynamic gestures in American Sign Language (ASL). Their results showed a significant improvement in the system's ability to anticipate gestures before their completion, representing an important step toward building more interactive and faster intelligent translators.

The field of Arabic Sign Language recognition has witnessed notable development in recent years, driven by advances in Computer Vision and Deep Learning techniques, which has led to improved methods for processing and analyzing gestures in real time. Recent studies have focused on building accurate and diverse datasets, with approaches ranging from the use of static images to temporal video clips in order to expand motion and visual representation. For example, [13] used the ArASL dataset while applying normalization and noise reduction operations, whereas the study by [14] relied on synthetic generation techniques (Augmentation) to ensure class balance. Other studies, such as [15],[16] focused on improving visual representation through reflection, scaling, and lighting adjustment to enhance model stability during training.

Dynamic sign video processing approaches have also emerged, such as the study by [17], who used frames extracted from visually filtered videos to train a Transformer Vision-Based model. Subsequent studies, such as [18] employed similar methods to capture fine-grained temporal patterns, while the study by [19] applied edge detection techniques to extract the fundamental features of the hand and fingers from complex backgrounds. In another context, the study by [20] integrated multiple data sources to enhance the comprehensiveness of visual representation.

At the algorithmic level, Convolutional Neural Networks (CNNs) have formed the cornerstone of many models due to their ability to extract fine-grained features of hand shape. [21] designed a CNN to classify Arabic alphabet letters with an accuracy exceeding 97%, while [22] developed a practical application for recognizing Saudi Sign Language using CNN and Transfer Learning techniques. Likewise, [23] presented a CNN-based model implemented using PyTorch with optimization algorithms such as Adam and RMSProp, achieving an accuracy of 97.9%.

Hybrid and transformer-based models have also emerged, combining the spatial capability of Convolutional Neural Networks (CNNs) with the temporal capability of recurrent and transformer networks. [24] provided an analytical review confirming the superiority of CNN-LSTM models in the representation of motion sequences. [23] developed a model that integrates CNN and LSTM with high generalization capability. Furthermore, the study by [19] in Chinese Sign Language highlighted the effectiveness of transformer-based models such as YOLO, CapsNet, and CNN-D. Other studies integrated Transfer Learning approaches with ResNet and EfficientNet, achieving accuracies exceeding 98%.

Collectively, these studies highlight a gradual scientific trajectory toward building advanced hybrid models that employ spatial and temporal analysis, progressing from the recognition of individual letters and isolated signs to understanding continuous signed sentences. Despite the significant progress, there remains a clear need to develop systems capable of supporting real-time predictive translation, which this project seeks to achieve by integrating computer vision techniques, temporal analysis, and motion prediction to provide faster, more accurate, and more realistic translation.

III. THEORIES USED IN BUILDING THE MODEL

The proposed research relies on a set of theories, each of which is directed to serve a specific component of the model, thereby contributing to the construction of an integrated system capable of recognizing Arabic Sign Language gestures with accuracy and efficiency. The following provides a detailed explanation of how each theory is employed within the project:

A. Predictive Processing Theory

This theory assumes that the brain predicts the upcoming movement before it occurs and continuously adjusts its expectations. The model benefits from this principle to predict the next sign and provide near real-time translation during motion execution.

B. Visual Attention Theory

This theory focuses on selecting the most important visual elements within the scene, such as hand shape and finger orientation. It helps enhance the model's ability to ignore noise and focus solely on the essential sign.

C. Sequential Learning Theory

This theory assumes that understanding a sign depends on analyzing a sequence of movements over time rather than treating each movement as an isolated action. This enables the system to interpret complete signed sentences instead of individual gestures.

D. Convolutional Neural Networks Theory (CNN Theory)

This theory is used to extract visual features from sign images. By integrating the MobileNetV2 architecture, which is fully based on convolution and pooling layers, the model analyzes hand shape, finger orientation, and fine-grained visual details within the image.

E. Transfer Learning Theory

This theory contributed to accelerating the training process, improving accuracy, and reducing the need for large-scale datasets, as the project leveraged pre-stored knowledge within the network and adapted it to suit Arabic Sign Language gestures.

F. Transformer Theory

This theory contributes to processing complex signs and sequential sentences and represents a scientific foundation for any future development toward building a real-time sentence-level translation system rather than one limited to individual letters.

G. Temporal Learning Theory in Deep Learning

This theory was integrated as it represents a cornerstone for the future development of the system toward translating dynamic signs and signed sentences that require understanding motion over time.

H. Explainable AI Theory

Although the project does not include a direct implementation of explain ability tools such as LIME or SHAP, this theory represents an important framework to ensure model transparency and reliability, especially when dealing with sensitive groups such as the deaf and hard of hearing.

I. Computer Vision Framework

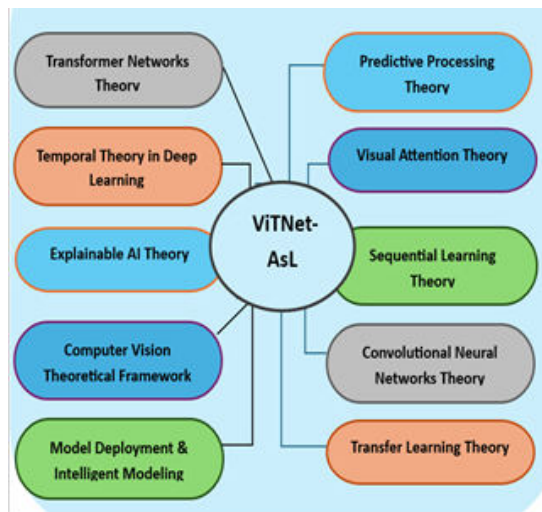
This framework was incorporated for image preprocessing (cleaning, resizing, normalization), color handling and transformations using OpenCV, data augmentation, and visual feature extraction through MobileNetV2, in addition to employing Mediapipe for hand landmark extraction when needed.

J. Model Deployment & Optimization Framework

This framework was introduced to provide the project with real practical value. A set of deployment, acceleration, and optimization processes was implemented, including converting the model from PyTorch to ONNX, using onnxruntime to test it across different environments, simplifying the model using onnxsim, accelerating inference using Optimum/Intel, and supporting processing through GPU units.

IV. ALIGNMENT OF DESIGN STAGES WITH THE APPLIED THEORIES

The proposed ViTNet-ASL model consists of three interconnected main layers. It starts with the first layer, which is preprocessing, followed by the second layer responsible for feature extraction and fusion, and ends with the third layer, which is gesture classification and text conversion, as illustrated in the following table that presents the core component, functional role, and the theories used for each layer.



The Main Theories on Which the Model is Based

Fig 1: Model Architecture

V. STAGES OF DESIGNING THE PROPOSED MODEL

The following figure illustrates the workflow of the proposed model such as [23],[22] starting from capturing the image through the camera, then passing it through preprocessing and feature extraction stages, and finally reaching the classification stage and generating Arabic text in real time.

The proposed ASL-ViTNet model consists of three interconnected main layers to achieve accurate translation of Arabic Sign Language.

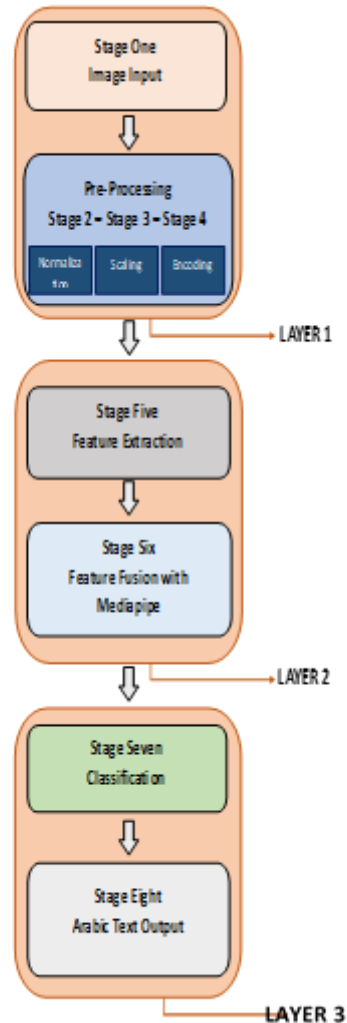


Fig 2: Stages of Designing

A. Layer 1

In this layer, the image is input into the system, and the preparation process begins through natural resizing, quality enhancement, and removal of any noise that may affect the accuracy of analysis. This step aims to highlight image features and ensure their readiness for visual feature extraction.

This stage is based on the Computer Vision Framework, which defines the fundamental principles of image processing, such as organizing visual data and determining enhancement steps to prepare it for subsequent models. In addition, Convolutional Neural Networks Theory is utilized to understand basic visual features, as this theory explains how convolutional networks are able to capture edges, contrasts, and structural patterns of the image, helping to form an appropriate initial visual representation. This stage is further supported by data augmentation techniques, due to their role in generating diverse versions of the image through rotation, cropping, or slight lighting modifications, in order to ensure balanced visual representation, enhance the model’s generalization capability, and reduce the risk of overfitting.

B. Layer 2

After data preparation, the images move to the stage of extracting fine-grained visual features using the MobileNetV2 model based on Convolutional Neural Networks Theory. The role of this theory here is to enable the model to identify important visual patterns and reduce dimensionality without losing essential information, in addition to improving extraction efficiency in a manner suitable for resource-constrained systems.

At this stage, the spatial structures of the image are analyzed, and then the data are fed into the ViT model to analyze spatial relationships according to Transformer Theory. This theory operates by using a self-attention mechanism to analyze correlations between image components and distinguish the most important parts. Mediapipe units are also used as geometric filters to support discrimination accuracy. This stage further integrates principles of Transfer Learning Theory to improve performance with fewer resources, as it allows the reuse of previously acquired knowledge from models trained on large-scale datasets. In addition, Sequential Learning Theory is leveraged when dealing with sequential patterns within the sign, to enhance the system's ability to understand temporal and spatial relationships that may appear within image data or its sequential representation.

C. Layer 3

The extracted features are sent to the classification layer, which determines the corresponding sign or letter. The outputs are then translated into Arabic text through a linguistic module operating in real time. This stage relies on Transformer Theory to supports high-accuracy classification, as this theory enables the system to understand the relationships between visual features and linguistic output through the attention mechanism, thereby enhancing the accuracy of the final decision.

At the text generation level, Sequential Learning Theory is used to interpret the motion sequence, as it allows the system to construct coherent sentences by predicting the next word or letter based on the preceding context.

In addition, Predictive Processing Theory contributes to enhancing the system's ability to anticipate the upcoming linguistic output, which improves text fluency and quality.

Finally, Explainable Artificial Intelligence (Explainable AI Theory) is adopted to interpret the decision-making mechanism within the system and clarify the reasons that led to selecting a specific sign or the resulting text. This provides necessary transparency and enables model improvement and documentation of its decisions in a scientific manner which is showed on figure below.

VI. RESULTS AND EVALUATION

Figure 3below: Login Interface of the IsharaTECH Application. The welcome screen allows the user to start the translation process by pressing the camera button and represents the main entry point for interaction with the system. Another Screen below is Real-Time Translation ModeDisplays the actual translation screen, where the user is asked to place their hand inside the frame to begin live sign recognition using artificial intelligence.



Fig 3. : Login Interface of the Ishara

A. Training and Testing Results

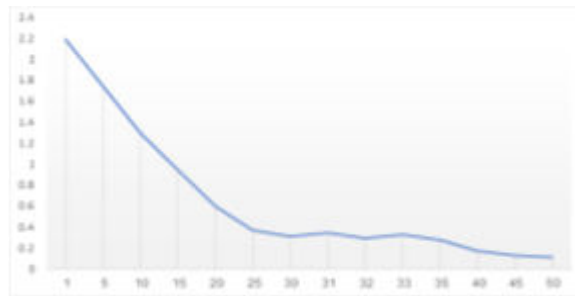
The model was trained using 19,840 images over 50 training epochs, with a batch size of 20 images per batch, while applying an early stopping algorithm to avoid overfitting. The results showed a continuous decrease in the loss value and a gradual increase in training accuracy until reaching a stable state. At the testing level, the model was evaluated using



4,960 test images and achieved an overall classification accuracy of 94.8%, with a Precision of 95.3%, Recall of 93.9%, and an F1-Score of 94.6%.

B. Test Loss Curve Across Evaluation Cycles

Figure 5 illustrates the gradual decrease in validation loss across 50 evaluation cycles. The loss started at a value of 2.20 and steadily declined to reach 0.12. Some minor fluctuations are observed around cycle 30, followed by a clear stabilization phase after cycle 40, indicating model maturity and stable learning behavior.



C. Model Performance Under Different Testing Conditions

Table I, below presents a comparison of the model’s performance on specialized test sets representing different real-world conditions.

Table I: Performance under Different Testing Conditions

Test Set	# of Samples	Accuracy (%)
Baseline Data (Ideal)	2,480	97.9
Low Lighting	620	90.5
Complex Backgrounds	620	92.8
Non-Standard Camera Angles	620	94.1

D. Model Performance

Table II Performance Comparison with Previous Models

Model	Accuracy (%)	Main Technique	Added Advantage
CNN (Abdelghfar et al., 2024)	91.2 %	Convolutional Neural Network	Manual feature extraction
ResNet-50 (Balat et al., 2024)	92.7 %	Transfer Learning	ImageNet pretrained weights
ViT-Base (Baghdadi et al., 2024)	93.5 %	Vision Transformer	Advanced attention mechanism
Proposed ASL-ViTNet	94.8 %	CNN + ViT + Mediapipe	Fusion of hand-crafted geometric and visual features



Table III Performance Comparison

Metric	Achieved Value (%)	Interpretation
Accuracy	94.8%	High overall accuracy in classifying signs.
Precision	95.3%	The model's ability to avoid incorrect classifications.
Recall	93.9%	Efficiency in detecting correct signs.
F1-Score	94.6%	Balance between precision and recall.

D. Discussion

The results of the ViTNet-ASL model demonstrate a high level of performance that reflects the effectiveness of the hybrid architecture combining MobileNetV2, Vision Transformer (ViT), and Mediapipe, as the model achieved a testing accuracy of 94.8% with balanced performance indicators. These results can be explained by several interrelated factors. On one hand, the use of MobileNetV2 in the spatial feature extraction stage contributed to providing lightweight and precise visual representations, which improved the model's ability to distinguish hand shapes and finger positions even under different lighting conditions. In addition, the integration of geometric features extracted from Mediapipe played an important role in reducing confusion between visually similar signs, as it provided additional information about joints and motion direction, thereby supporting model stability during classification.

On the other hand, the use of Vision Transformer (ViT) helped enhance the understanding of deep spatial patterns within the sign through the self-attention mechanism, which explains the model's high accuracy in recognizing most letters. The transformer-based architecture also supported the model's ability to interpret the sign as an integrated visual entity rather than analyzing each component separately, which contributed to improving classification accuracy and balancing performance metrics.

When analyzing the training behavior, it can be observed that the model exhibited a steady decrease in loss values and a continuous improvement in accuracy across epochs. This is attributed to the application of structured preprocessing, including normalization and data augmentation, which enhanced the model's generalization capability and prevented overfitting. The Early Stopping algorithm also helped preserve the optimal model without allowing training to continue into stages where efficiency might decline.

The observed confusion between the letters (S) and (Z) is due to the high similarity in the motion shape of the sign, particularly in terms of finger positioning and arrangement, which is a common challenge in most models that rely on static images. Nevertheless, classes such as (A) and (M) achieved an accuracy of 100%, owing to the clarity of their signs and their clear visual distinction compared to other letters, which made the discrimination process easier for the model.

Overall, these results indicate that the integration of visual and geometric feature analysis, along with the utilization of attention mechanisms and deep analysis through ViT, played a decisive role in achieving accurate, stable, and reliable recognition of Arabic Sign Language letters in achieving high performance. These findings indicate that the model is not only accurate, but also capable of generalization and handling sign variability, which confirms the effectiveness of the hybrid architecture adopted in this project.

VII. CONCLUSION

The proposed ASL-ViTNet model has demonstrated high effectiveness in recognizing Arabic Sign Language gestures and translating them into written text with an accuracy of 94.8%.

The integration of Vision Transformer, Mediapipe, and MobileNetV2 provided a balanced analysis between deep visual representations and the geometric properties of the hand, making the model suitable as an assistive intelligent communication technology for the deaf and hard of hearing.

REFERENCES

- [1] A. Almjally and W. S. Almkadi, "Deep computer vision with artificial intelligence-based sign language recognition to assist hearing and speech-impaired individuals," *Scientific Reports*, vol. 15, no. 1, pp. 1–14, 2025, doi: 10.1038/s41598-025-09106-8.
- [2] S. A. A. Shah, A. Bennamoun, F. Boussaid, and M. El-Hajj, "A review on vision-based hand gesture recognition systems," *IEEE Access*, vol. 8, pp. 131920–131944, 2020, doi: 10.1109/ACCESS.2020.3008223.
- [3] R. Duwairi and Z. A. Halloush, "Automatic recognition of Arabic alphabets sign language using deep learning," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2996–3004, Jun. 2022, doi: 10.11591/ijece.v12i3.pp2996-3004.
- [4] H. El-Badawy, A. M. Hassan, and S. Aly, "Arabic sign language recognition: Challenges and future directions," *J. Comput. Commun.*, vol. 12, no. 4, pp. 45–58, 2024.
- [5] Y. Zhou, G. Chen, and J. Wang, "Spatial-temporal modeling and early prediction for real-time sign language recognition," *IEEE Access*, vol. 11, pp. 98765–98778, 2023.
- [6] S. B. Jugunta et al., "Exploring the Insights of Bat Algorithm-Driven XGB-RNN (BARXG) for Optimal Fetal Health Classification in Pregnancy Monitoring," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 11, pp. 731–741, 2023, doi: 10.14569/IJACSA.2023.0141174.
- [7] A. O. Alsayed, N. A. Ismail, L. Hasan, and F. Embarak, "A Comprehensive Review of Modern Methods to Improve Diabetes Self-Care Management Systems," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 9, pp. 182–203, 2023, doi: 10.14569/IJACSA.2023.0140920.
- [8] A. Elhadad et al., "Improved healthcare diagnosis accuracy through the application of deep learning techniques in medical transcription for disease identification," *Alexandria Eng. J.*, vol. 123, pp. 112–123, Jun. 2025, doi: 10.1016/J.AEJ.2025.03.047.
- [9] F. Embarak, N. A. Ismail, and S. Othman, "The impact of online social community platforms on the elderly daily activities: A systematic literature review," 2020, Accessed: Jul. 25, 2025.
- [10] A. O. Alsayed, N. A. Ismail, L. Hasan, M. Binsawad, and F. Embarak, "Leveraging a hybrid convolutional gated recursive diabetes prediction and severity grading model through a mobile app," *PeerJ Comput. Sci.*, vol. 11, p. e2642, Feb. 2025, doi: 10.7717/PEERJ-CS.2642/SUPP-3.
- [11] Y. Cheng, J. Hu, and X. Yang, "Early prediction of continuous Chinese sign language gestures using deep sequential neural networks," *IEEE Access*, vol. 13, pp. 84521–84533, 2025.
- [12] M. Barham-Al Muhammad, A. Alsharif, and N. Alotaibi, "Arabic sign language recognition using ArASL dataset with preprocessing and noise reduction," *Multimedia Tools and Applications*, vol. 83, no. 9, pp. 24317–24335, 2024.
- [13] E. R. Kharoua, "Data augmentation-based deep learning framework for Arabic sign language recognition," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, pp. 321–330, 2023.
- [14] F. Embarak, N. Azman Ismail, O. R. Shahin, and R. N. Alabdali, "Design of autonomous online social community architecture for older adults(pdf)," *Comput. Electr. Eng.*, vol. 100, May 2022, doi: 10.1016/j.compeleceng.2022.107900.
- [15] F. Embarak, N. Azman Ismail, and O. R. Shahin, "Design of autonomous online social community architecture for older adults," *Comput. Electr. Eng.*, vol. 101, p. 107900, Jun. 2022, doi: 10.1016/J.COMPELECENG.2022.107900.
- [16] A. N. Baghdadi, M. Al-Dosari, and S. Benhaddi, "Continuous Arabic sign language recognition from video sequences using deep temporal modeling," *IEEE Access*, vol. 12, pp. 158901–158913, 2024.
- [17] M. F. Almufareh, A. Alqahtani, and S. Alshammari, "Vision-based transformer model for sign language recognition using temporally filtered video frames," *IEEE Access*, vol. 13, pp. 112345–112357, 2025.
- [18] A. Khuzayem-Al, M. Alsharif, and H. Almutairi, "Saudi sign language recognition using CNN and transfer learning," *Computers & Electrical Engineering*, vol. 113, pp. 108934, 2024.
- [19] A. Al-Rifaai, A. Abdulsalam, and M. Al-Sabah, "Deep CNN-based Arabic sign language recognition using PyTorch with optimization algorithms," *IEEE Access*, vol. 12, pp. 174220–174231, 2024.
- [20] A. Asmaa, A.-H. Marwa, A. Alaa, E. Farhat, And E. Abdulhamid, "Cultivating Independent Learners: Designing An Innovative Interactive Model To Enhance Self-Directed Learning And Autonomy In Elementary Education," *Int. J. Innov. Res. Sci. Eng. Technol.*, Vol. 14, No. 1, P. 75, Jan. 2025, Accessed: Mar. 18, 2025.
- [21] A. O. Alsayed, N. A. Ismail, L. Hasan, M. Binsawad, and F. Embarak, "Leveraging a hybrid convolutional gated recursive diabetes prediction and severity grading model through a mobile app(PDF)," *PeerJ Comput. Sci.*, vol. 11, p. e2642, 2025, doi: 10.7717/PEERJ-CS.2642/SUPP-3.
- [22] Asmaa, Aljilany, A.-H. Marwa, A. Alaa, E. Farhat, and E. Abdulhamid, "Cultivating Independent Learners: Designing an Innovative Interactive Model to Enhance Self-Directed Learning and Autonomy in Elementary Education," *International Journal of Innovative Research in Science Engineering and Technology*. Accessed: Jan. 28, 2025.



- [23] F. Embarak Et Al., “Online Social Interaction Community Model For Unmet Needs Of Older People,” J. Theor. Appl. Inf. Technol., Vol. 101, P. 22, 2023, [Online]. Available: Www.Jatit.Org
- [24] F. Embarak, A. Eltheeb, A. O. Alsayed, And D. N. P. Devadhas, “Designing Personalized Assistive Technologies For Elderly Users In Human-Computer Interaction For Haptic Recognition Using Gnn,” J. Theor. Appl. Inf. Technol., 2025, Accessed: Dec. 29, 2025.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details