Journal of Engineering and Technology

ISSN 2180-3811 eISSN 2289-814X https://jet.utem.edu.my/jet/index

DOI: https://doi.org/10.54554/jet.2025.16.2.012

CAMERA-BASED REAL-TIME QURANIC SIGN LANGUAGE DETECTION SYSTEM USING LSTM DEEP LEARNING SEQUENCES

M. U. Ahmad Adli[1], N. H. Zainun Anuar[1], K. Abdulrahim*[1] and K. N. Zainul Ariffin[1]

[1] Department of Electronic and Electrical Engineering, Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia, 71800, Nilai, Malaysia.

*corresponding: khairiabdulrahim@usim.edu.my

Article history:

Received Date: 30 July 2024

Revised Date: 9 April 2025

Accepted Date: 18 June 2025

Keywords: LSTM, Quranic Sign Language, Communication, Deep Learning, Disability

Abstract— Effective communication, vital for expressing emotions, ideas, and resolving conflicts, depends on various forms of language, including written symbols, gestures, and vocalizations. While a shared language often facilitates successful communication, a significant challenge arises between individuals who rely on sign language due to speech impairments and those who use spoken languages. This gap creates barriers to mutual understanding. This study addresses the issue by implementing a proficient deep learning model designed to predict Quranic sign language, aiming to bridge the communication divide between speech-impaired and non-speech-impaired individuals. The research employed a Long Short-Term Memory (LSTM) model, and the results demonstrated that the LSTM model achieved superior performance in recognizing and interpreting Quranic sign language, highlighting its potential as a tool to enhance inclusivity within the community.

I. Introduction

Persons with Disabilities (PWDs) are categorized into seven groups: hearing, visual, speech, physical, learning, mental, and multiple disabilities [1]. Communication between individuals with disabilities and those without can often be challenging for various reasons. As of 2017, the Malaysian Department of Social Welfare recorded 453,238 registered PWDs, with the following distribution: 35.2% with physical disabilities, 34.8% with learning disabilities, 8.9% with visual impairments, 8.3% with mental disabilities, 4.7% with multiple disabilities, and 0.5% with speech-related disabilities [2].

When conventional communication methods, such as speaking, are difficult, people with disabilities often adopt alternative means of expression. Sign language, recognized as a visual language, is one of the most effective tools for individuals with speech and hearing impairments. Languages like American Sign Language (ASL) and British Sign Language (BSL) are widely utilized across various communities.

For the deaf and hard-of-hearing Muslim community, a unique adaptation known as Quranic Sign Language has been developed. This specialized form of sign language aims to facilitate the understanding and recitation of Quranic verses and teachings through specific gestures that capture the rich and nuanced meanings of the Quran.

Unfortunately, a lack of accommodations, such as interpretation services in local mosques, often limits the participation of deaf individuals in Islamic worship [3].

This project seeks to address the challenges of translating Quranic sign language by leveraging deep learning techniques, specifically Long Short-Term Memory (LSTM) networks, combined with camera-based finger movement detection. The objective is to develop an automated system capable of recognizing Quranic sign language gestures and translating them into a format understandable to individuals unfamiliar with sign language. This effort is particularly significant given the distinct nature of Quranic sign language alphabets compared to other sign languages, as illustrated in Figure 1.

Figure. 1: Quran Sign Language Alphabets

II. Literature Review

Many researchers are exploring how technology can support individuals who are deaf or hard of hearing by advancing sign language recognition.

Utilizing computer vision and machine learning, these studies aim to understand and interpret the hand movements and gestures inherent in sign language. For instance, research on Posture Recognition in American Sign Language Using Neural Networks [4] revealed that accurately identifying hand shapes is crucial for comprehending sign language. Such efforts are directed toward improving communication and accessibility for sign language users.

One notable study presented a vision-based sign language recognition system that combined a multi-view CNN architecture with depth features [5]. This system demonstrated the feasibility of camera-based methods in real-world scenarios by successfully recognizing Indian Sign Language (ISL) gestures. Similarly, To et al. [6] developed a camera-based American Sign Language (ASL) translation system that utilized depth and skeleton data. By employing a two-stream deep neural network, the study achieved accurate, real-time ASL recognition and translation, highlighting the potential of camera-based approaches for real-time sign language detection.

Deep learning has proven to be an effective tool for sign language recognition, significantly advancing the accuracy and efficiency of gesture detection. Luo et al. [7] proposed an LSTM-based model for Chinese Sign Language recognition, demonstrating the ability of LSTM networks to capture temporal features in sign language gestures. Likewise, Dhulipala et al. [8] employed LSTM networks for continuous sign language recognition, underscoring LSTM's suitability for sequential data processing.

The combination of camera-based systems and LSTM networks offers a robust framework for detecting and interpreting uninterrupted sign language gestures. Camera-based systems excel at real-time gesture capture, while LSTM networks effectively handle sequential data, enabling precise interpretation of signing sequences. Dias et al. [9]

explored this synergy in Brazilian Sign Language (Libras), using CNN and LSTM to recognize a diverse set of gestures with high accuracy.

This study aims to develop a camera-based real-time system using LSTM deep learning for detecting Quranic sign language. Motivated by the need to enhance accessibility for individuals with hearing impairments in Quranic studies, the research addresses a significant gap by focusing on the unique gestures and religious symbols specific to Quranic sign language. By leveraging the strengths of camera-based systems and LSTM networks, the proposed system seeks to facilitate accurate and efficient interpretation of Quranic sign language in real-world applications.

III. Methodology

A. Software Requirements

The system operates on Windows 7 and above, using Jupyter Notebook as the platform, with Python as the programming language. Several libraries are employed to support the project. OpenCV, an open-source library, is utilized for its real-time application capabilities and computational efficiency when handling large data volumes, making it ideal for object detection and video processing to recognize signs and gestures in images and videos. TensorFlow, another open-source AI library, enables the construction of models using data-flow graphs, allowing the creation of large-scale neural networks with multiple layers. Keras, integrated with TensorFlow, is used to develop neural networks with LSTM layers for managing sequences of key points. MediaPipe provides a framework for constructing machine learning pipelines to process time-series data such as video and audio, specifically using MediaPipe Holistic to detect landmarks for the face, hands, and body poses simultaneously. Finally, Scikit-learn (sklearn) is employed to assess system performance, utilizing built-in metrics and confusion matrices for accuracy evaluations.

B. Hardware Requirements

The system requires a camera with good quality, at least 3MP, and a minimum of 8GB RAM for smooth operation. A GPU with 4GB dedicated memory is essential, along with a processor equivalent to or higher than the Radeon R5. Storage needs include at least 10GB of HDD space, and a 15inch or 17inch color monitor is recommended for optimal display.

C. Project Design

As illustrated in Figure 2, the architecture of the project begins with capturing a video of the person using OpenCV, which serves as the input. Next, the data is collected using MediaPipe Holistic, which detects keypoints related to the face, pose, and hand landmarks. These keypoints are stored in frames of a video format, with the data pushed into a NumPy array for further processing. The system is then trained using a Long Short-Term Memory (LSTM) deep learning model, consisting of three LSTM layers followed by three Dense layers [10]. The model is trained for 30 epochs with a batch size of 128 using the extracted dataset. The training process aims to minimize the loss through categorical cross-entropy and is optimized using the Adam optimizer. Finally, after building the neural network, real-time sign language recognition is performed with OpenCV, where the recognized gestures are displayed as text within a highlighted section.

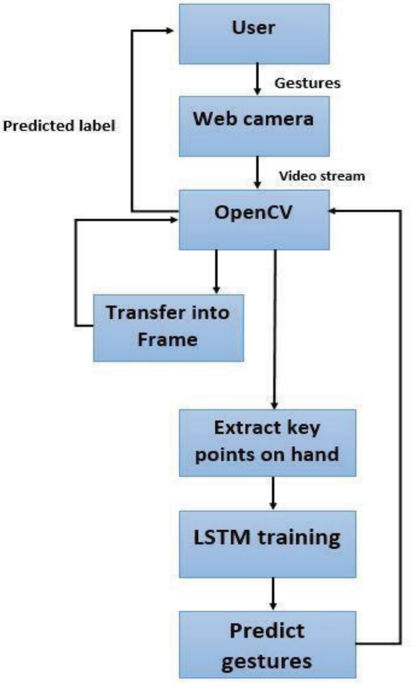


Figure 2: Project System Flowchart

D. Long-Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture developed to address the limitations of traditional RNNs in handling long-term dependencies in sequential data. LSTMs are particularly effective for tasks involving sequences, such as natural language processing and speech recognition. Unlike other sequence learning models, such as hidden Markov models or traditional RNNs, LSTMs have memory cells that can store and retrieve information over extended periods. This ability allows LSTMs to capture and remember context and dependencies in sequential data, making them highly efficient for tasks where understanding the temporal relationships between elements is essential. Additionally, LSTMs offer an advantage over other models due to their capability to manage relative intensity gaps, which further enhances their performance in sequential tasks [11].

E. Data Description

In this system, the required data were about 28 signs used in Quranic sign language. However, only 3 signs will be used in this project, as shown in Figure 3 because of the limitation on the device. The signs were based on three categories, namely, hand gesture signs, pose signs, and facial expressions, or on a combination of facial expressions with either hand gestures or pose signs.

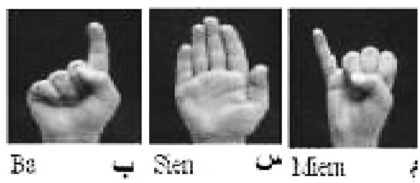


Figure 3: Numeric Signals Chosen for This Project

F. LSTM Model Specification

The main goal of this experiment was to implement an LSTM model based on to predict Quranic sign language using multiple frames and to predict the action being demonstrated in real-time. The step-by-step procedure to construct this system is as follows:

i) Installing and importing dependencies: Set up the necessary libraries and tools required for the project

ii) Key points using MediaPipe holistic: Use MediaPipe Holistic to detect key points of the face, hands, and body for accurate gesture recognition

iii) Extract Key Points: Collect the key points from the video frames to be used in the model

iv) Setup folders for data collection: Organize the project structure by creating folders to store the dataset

v) Collect Keypoint sequences: Gather sequences of key points over multiple frames to form the dataset for training

vi) Preprocess data and create labels: Clean and preprocess the data, and assign appropriate labels to each sequence for training

vii) Build and Train an LSTM deep learning model: Construct an LSTM model to handle sequential data and train it on the prepared dataset

viii) Make sign language predictions: Use the trained model to predict sign language gestures from new input

ix) Save model weights: Save the trained model weights for future use and deployment

x) Evaluation using confusion matrix and accuracy score: Assess the model's performance by evaluating it with a confusion matrix and calculating the accuracy score

xi) Test in real-time: Test the system in real-time to verify its effectiveness in predicting Quranic sign language gestures

IV. Results and Discussion

A. Keypoint Extraction Result

Figure 4 illustrates the process of keypoint extraction, a crucial step for detecting and localizing facial and hand landmarks in images or video frames. This process is essential for recognizing gestures and expressions. In this project, facial keypoints are identified to represent features such as the eyes, nose, mouth, and facial outline, while hand keypoints correspond to joints and fingertips. These keypoints form the basis for analyzing and interpreting human gestures and expressions in visual data.

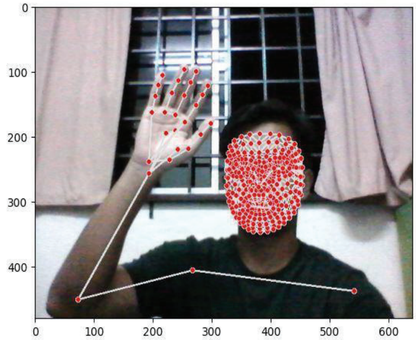
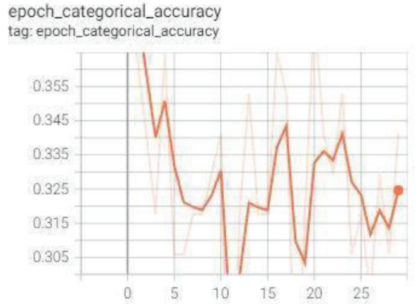


Figure 3: Keypoints Extraction for Face and Hand

B. Epoch Accuracy and Lost Result

The current training graphs from Figure 5 show that the model is encountering difficulties in achieving high performance, as indicated by the erratic accuracy values and the sharp initial increase followed by a gradual decrease in loss. These fluctuations suggest that the model is struggling to consistently learn from the data, possibly due to factors like an inappropriate learning rate, insufficient data quality, or the need for further optimization of the model architecture. However, the stabilization observed towards the end of the training process indicates that the model has reached a certain level of learning, even though it has not yet achieved the desired performance. Identifying these issues early on is advantageous, as it provides a clear direction for necessary adjustments and improvements to enhance the model's performance



(a)

(b)

Figure 5: Epoch Accuracy and Loss Graph for 30 Epoch

C. Realtime Test Result

The real-time results shown in Figure 6, 7 and 8. The results show that the model successfully identifies the signs 'ba', 'sein', and 'meim'. During LSTM training, the model learns from a dataset of sign language gestures, with the goal of accurately distinguishing between different signs. However, challenges arise due to variability in how the signs are executed and environmental factors that can affect recognition accuracy. To improve performance, it is crucial to refine the dataset by including more diverse examples, optimize the LSTM architecture, and enhance the model's robustness. These adjustments will help ensure reliable and consistent real-time sign language recognition.

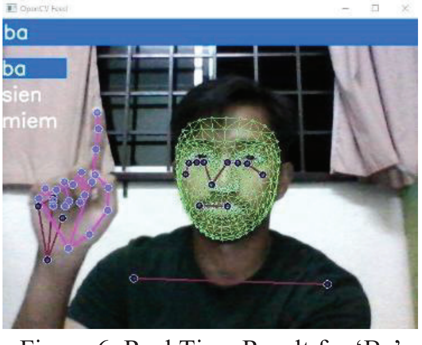


Figure 6: Real Time Result for ‘Ba’ Sign

Figure 7: Real Time Result for ‘Sein’ Sign

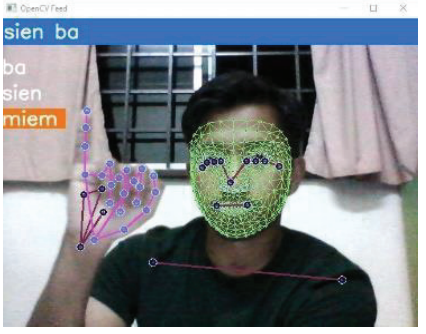


Figure 8: Real Time Result for ‘Meim’ Sign

D. Percentage Accuracy of Sign for 10s

Figure 9 shows the percentage of data when the user performs the 'ba' sign for 10s. The result for the 'ba' sign within 10s is not 100% accurate because the LSTM model may misinterpret other data points due to overlapping features or similarities with other signs. This can cause the model to mistakenly read data from other movements or positions, reducing overall accuracy. The highest percentage for each sign indicates that the collected results are accurate, with 'ba' at 66.7%, followed by 'sein' at 33.33%, and 'meim' at 22%.

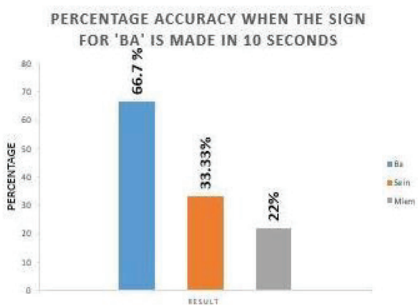


Figure 4: Percentage Accuracy for 'ba' Sign Made in 10s

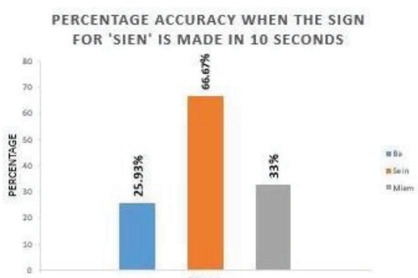


Figure 10: Percentage Accuracy for 'sein' Sign Made in 10s

Figure 10 similarly shows the percentages for the sign 'sein' over 10s. Within those 10s, it indicates readings of 'meim', but not very accurately due to the factors mentioned earlier. The highest percentage for each sign indicates that the collected results are accurate, with 'sein' at 66.7%, followed by 'meim' at 33.33%, and 'ba' at 25.93%.

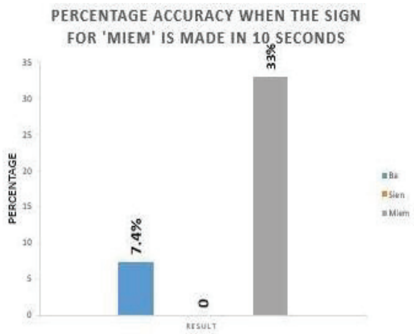


Figure 5: Percentage Accuracy for 'meim' Sign Made in 10s

Figure 11 presents the percentages for the sign "meim" recorded over a span of 10 seconds. During this period, the data reflects readings of "meim," albeit with limited accuracy due to various influencing factors. The highest percentage recorded for any sign indicates the reliability of the results, with "meim" and "ba" both registering 33% and 33.33%, respectively. However, the percentage for "sein" is observed to be 0%, which may be attributed to the similarity between the "meim" and "ba" signs, leading to potential misclassification.

V. Conclusion

In conclusion, the developed system successfully achieved its primary goal of predicting Quranic sign language in real-time using an LSTM model. This approach aims to facilitate a more accessible understanding of Quranic teachings for individuals with hearing difficulties. The methodology not only aligns with current trends in intelligent learning but also offers a promising solution to practical challenges, ultimately promoting inclusivity and accessibility in Quranic education. The positive evaluation outcomes, combined with considerations for practical implementation, confirm the system's potential for real-world applications in interpreting Quranic sign language gestures. As technology continues to advance, this system represents a significant step toward bridging communication gaps and enhancing accessibility for individuals using Quranic sign language.

VI. Acknowledgement

The research leading to this paper was partly supported by Universiti Sains Islam Malaysia under the grant number PPPI/USIM/FKAB/USIM/16124.

VII. References

[1] S. Jaffer, “Global group helping deaf Muslims embrace faith: Salt Lake Telegram,” *Deseret News*, 2014. [Online] Available: https://www.proquest.com/newspapers/global-group-helping-deaf-muslims-embrace-faith/docview/1552455607

[2] S. A. Muhammad *et al.*, “Assistive data glove for isolated static postures recognition in American Sign Language using neural network,” *Electronics*, vol. 12, no. 8, pp. 1904, 2023.

[3] S. Subburaj and S. Murugavalli, “Survey on sign language recognition in context of vision-based and deep learning,” *Measurement: Sensors*, vol. 23, p. 100385, Oct. 2022.

[4] Md. M. Islam, S. Siddiqua, and J. Afnan, “Real time hand gesture recognition using different algorithms based on American Sign Language,” *IEEE Xplore*, Feb. 2017. [Online] Available: https://ieeexplore.ieee.org/abstract/document/7890854

[5] H. Y. Luo and L.-h. Luo Yin, “Chinese sign language recognition based on two-stream

CNN and LSTM network," *International Journal of Advanced Networking and Applications*, vol. 14, no. 6, p. 5666, 2023.

[6] W. He, T. Huang, Y. Qiao, J. Yao, and X. Bai, "Sign language recognition using 3D convolutional neural networks," in *Proc. 23rd Int. Conf. Pattern Recognition (ICPR)*, 2016, pp. 433-438.

[7] T. Leal, U. Nunes, and J. Xavier, "CNN-based real-time hand gesture recognition for Brazilian Sign Language," in *Proc. 18th Symp. Virtual and Augmented Reality*, 2016.

[8] S. U. Amin, H. Altaheri, G. Muhammad, A. W. Abdul, and M. Alsulaiman, "Attention-inception and long-short-term memory-based electroencephalography classification for motor imagery tasks in rehabilitation," *IEEE Transactions on Industrial Informatics*.

[9] T. S. Dias, J. J. A. M. Junior, and S. F. Pichorim, "Comparison between handcraft feature extraction and methods based on recurrent neural network models for gesture recognition by instrumented gloves: A case for Brazilian sign language alphabet," *Biomedical Signal Processing and Control*, vol. 80, p. 104201, 2023.

[10] R. M. Abdulhamied, M. M. Nasr, and S. N. Abdulkader, "Real-time recognition of American sign language using long-short term memory neural network and hand detection," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 1, pp. 545-556, 2023.

[11] H. Alizadegan, B. Rashidi Malki, A. Radmehr, H. Karimi, and M. A. Ilani, "Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction," *Energy Exploration & Exploitation*, vol. 43, no. 1, pp. 281-301, 2025.