Mathematical Sciences and Informatics Journal
Vol. 3, No. 2, Nov 2022, pp. 65-79
http://www.mijuitmjournal.com DOI: 10.24191/mij.v3i2.19395

A Framework for Malaysian Sign Language Recognition using Deep Learning Initiatives

Imran Md Jelas
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch Tapah Campus, Perak, Malaysia
imran499@uitm.edu.my

Article Info
Article history:

Received Aug 28, 2022
Revised Oct 5, 2022
Accepted Oct 30, 2022

Keywords:

Malaysian Sign Language (MSL)
Sign Language Recognition (SLR)
Vision-Based SLR
Deep Learning
MediaPipe

ABSTRACT

The greatest challenge since the introduction of Malaysian Sign Language (MSL) occur when deaf and hard of hearing people try to communicate using MSL with person without disabilities who do not use MSL. To overcome this communication barrier, a substantial number of studies has been done to produce Malaysian Sign Language Recognition system. Given that MSL is a systematic nonverbal language that utilizes both manual signal and non-manual signal, the employment of a vision-based sign language recognition system denote applicability. A vision-based sign language recognition system utilizes hand direction, wrist orientation and joint angles detection on captured image to capture sign. In 2019, Google introduced MediaPipe a framework bestowed with face detection, hands detection and pose detection suitable for vision-based sign language recognition system. MediaPipe framework can simplified image processing stage tremendously which is crucial in vision-based sign language recognition system. Hence, the main objective of this paper is to develop a Framework for Malaysian Sign Language Recognition using Deep Learning. To achieve this objective, we propose a framework consisting of three main modules namely learning module, training module and detection module. The proposed framework will also be integrated with MediaPipe. Later, Long Short-Term Memory (LSTM) artificial neural network (ANN) is proposed as training algorithm in training module and prediction algorithm in detection module to be used for the development of the actual system based on this proposed framework initiative. LSTM, an ANN can recall both current data and past data. LTSM pertinent for vision-based sign language recognition system especially when continuous image is used as in the proposed framework.

Corresponding Author:
Imran Md Jelas
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch Tapah Campus, Perak, Malaysia
email: imran499@uitm.edu.my

1. Introduction

The advancement of computer science has becoming more embedded in human daily lives. This advancement is further extent as human try to leverage computer and technology to mimic the problem-solving skills and decision-making capabilities of the human mind through Artificial Intelligent (AI) [1]. AI works by implementation of Machine Learning (ML) algorithms. ML usually is human dependent thus sometimes requires human intervention to learn [1]. A scalable ML is known as Deep Learning (DL) [1]. DL automates feature extraction process in ML [2]. DL also eliminate if

PENERBIT PRESS
UNIVERSITI TEKNOLOGI MARA

not all, at least part of the manual human intervention required for the machine learning system to learn [3].

Notwithstanding the improvement in software development, this improvement has also expanded the Human Computer Interaction (HCI) hardware [4]. Innovation in HCI hardware has further improve human ability through Augmented Reality (AR) hardware or Virtual Reality (VR) hardware. AR and VR hardware improvement has produced substantial numbers of innovative hardware such as data glove, Kinect camera and leap motion controller[5]. This innovative hardware has been used extensively in sign language recognition as input acquisition devices. This can be observed in works by [5]–[13].

The progressive development in computer software through DL and innovative hardware for HCI can benefit individual with disability. People who are deaf and hard of hearing use sign language as their primary language [14]. In Malaysia, Malaysia Sign Language (MSL) was introduced as principal language of hearing impaired community in Malaysia [15]. The greatest challenge since the introduction of MSL occur when deaf and hard of hearing people try to communicate using MSL with persons without disabilities who do not use MSL [16]. Therefore, the aim of this study is to embark an initiative by proposing a framework for MSL recognition using DL towards overcoming this communication barrier. This will be made possible with the introduction of MediaPipe as new contribution in this proposed framework initiative.

2. Literature Review

This section discussed previous studies done to produce Malaysian Sign Language Recognition system. It also discussed MediaPipe framework that will be introduce and used in these initiatives.

2.1 Malaysian Sign Language Recognition

In order to overcome this communication barrier, a substantial number of studies has been done to produce Malaysian Sign Language Recognition system [4], [6], [15], [17]. MSL is a systematic nonverbal language that utilizes both manual signal and non-manual signal [6]. Manual signal is the finger movement and hand movement that passes on typical significance [6]. Non-manual signal includes head movement, body movement and body posture [6].

First, study in [4] and [6] focus on developing MSL recognition system by their own version of wearable or data glove. Research built data glove is not only costly to be build, but it also required abundant of sensor to capture signal. Both studies can be safely assumed as incomplete. The study in [4] only focuses on developing a wearable by placing a sensor on top and down part of wrist with another sensor slightly above wrist section to detect finger movement. Meanwhile, the study in [6] only focuses on developing data glove that will only able to capture finger movement. Even though combination of both studies might be able to capture manual signal, but both studies fall short in capturing non-manual signal.

However, study in [17] focuses on development of a dataset for MSL recognition by using Kinect camera. Kinect camera is suitable to be used as image acquisition device since Kinect camera enable the capture of RGB image, an IR depth image and skeleton joint which is suitable for vision-based sign language recognition system [17]. Kinect camera also able to capture audio but it is not use in vision-based sign language recognition system. But even with all this benefit, Kinect camera does not come as a standard in any computer system but needed to be bought separately thus increase the cost of overall sign language recognition system.

Finally, study in [15] focuses on the development of MSL recognition system by using a simple camera. It focusses more on image processing techniques to proceed with image classification. However, the system develop in this study shows poor centroid estimation when Linear Kalman Filter is used to detect overlapping during image pre-processing stage. Besides, the result also shows poor skin segmentation performance during image segmentation stage. Further, the system also facing feature vector issues during image extraction stage as it does not consider the finger movement.

Thus, this study aim is to propose improvement of image processing technique in [15]. This will be accomplished by capturing both manual signal and non-manual signal in MSL. The manual signal of finger movement and hand movement will be captured as in study [4] and [6] but without the additional cost of developing wearable or acquiring a data glove.

2.2 MediaPipe Framework

In 2019, Google introduce MediaPipe which serve as a framework for building machine learning pipeline to process continuous time-series data such as video or audio [18]–[20]. For video, MediaPipe enable landmark detection on stream video input from a simple camera or a webcam. This landmark detection includes face detection, right hand detection, left hand detection and pose detection. Upon detection of this landmark, MediaPipe can provide up to 468 keypoint for face detection, 21 keypoint for right hand detection, 21 keypoint for left hand detection and 33 keypoint for pose detection [18]–[20]. With this keypoint detection, MediaPipe may turn a simple camera or webcam into an input device resembling Kinect camera.

Not only MediaPipe is able to detect accurate keypoint based on each landmark, keypoint detected on each landmark can also be saved on continuous input image during image acquisition stage. Keypoint can be used to define region of interest. With MediaPipe, continuous input image gathered can almost directly be used for image classification with predictor algorithms. Using MediaPipe, input from a simple camera or webcam can be used for sign language recognition. Thus, researcher only need to focus on developing a scalable machine language and reduce as much as possible human intervention to create a Framework for Malaysian Sign Language Recognition using Deep Learning.

2.3 Image Acquisition Device

Image acquisition devices used by researcher to acquire sign language images for classification. The device can be categorised as simple devices (such as camera or webcam) or innovative HCI devices. In innovative HCI hardware aligns with AR and VR hardware improvement produce devices include data glove, Kinect camera and leap motion controller [5] but acquiring these devices will incur additional cost.

Camera or webcam is a basic camera that can capture still or moving images. While, Kinect camera can provide input like normal camera or webcam but with extra capability of gesture recognition and body skeletal detection [7]. Leap motion controller is a non-wearable device with optical sensors for gesture recognition [8], [9]. Data glove either proprietary product or research-built product is a wearable device for hand gesture recognition [6], [9]–[13].

In this paper, the proposed framework for Malaysian Sign Language Recognition using Deep Learning will use a simple webcam. Webcam is chosen as it typically comes as a standard hardware and it is able to provide satisfactory image that can be used during image segmentation stage and image extraction stage plus does not require image enhancements nor image restoration.

2.4 Image Processing Techniques

Image processing technique may briefly divide into three stage which is image pre-processing stage, image segmentation stage and image extraction stage.

2.4.1 Image Pre-Processing Stage

Implementation of image pre-processing stage is applied to acquire input image to remove unwanted noise and enhance the quality [14]. This can be accomplished by resizing, colour conversion, removing unwanted noise, or a combination of several of these techniques from the original acquire input image [14]. Image pre-processing stage can be classified into two processes. Image enhancement process is to improve input image quality. This process is crucial to better suited for human or machine analysis. Image enhancement process include Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), Contrast Limited Adaptive Histogram Equalization (CLAHE) and logarithmic transformation [14]. Image restoration process is to restore a degraded image caused by noise and blur. This process is greatly determined by the type of noise and corruption present in the image. Image restoration process includes median filter, mean filter, Gaussian filter, adaptive filter and wiener filter[14].

As previously discussed, webcam able to provide satisfactory image that can be used later during image segmentation stage and image extraction stage as shown in Figure 1. Thus, the images produce does not require image enhancements nor image restoration.

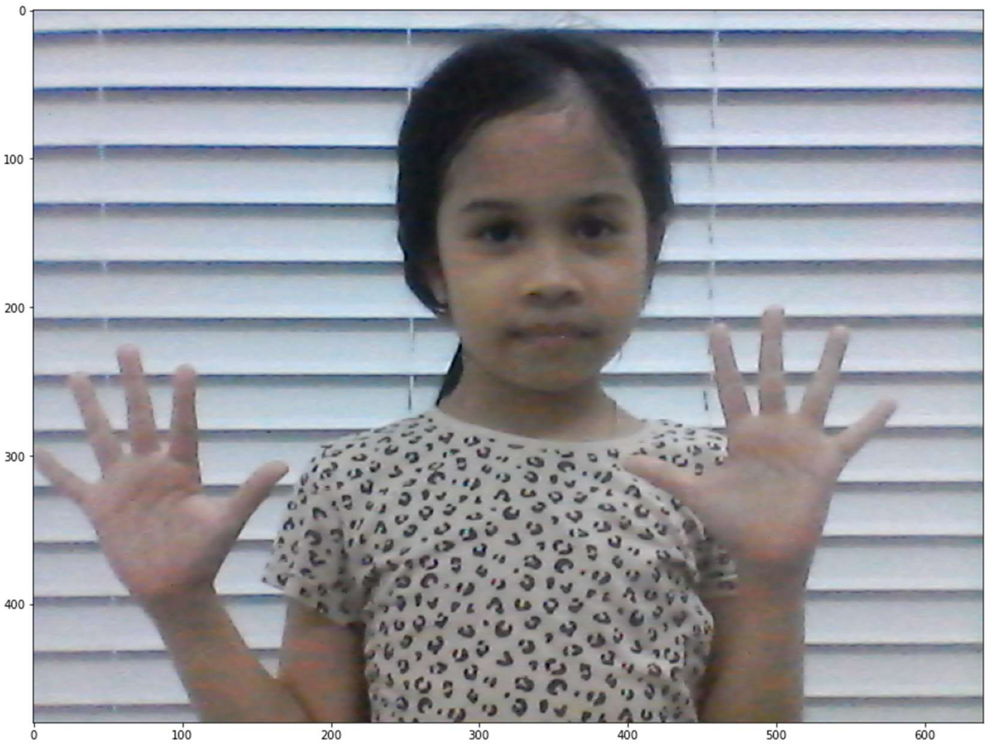

Figure 1. Satisfactory image acquisition with a simple laptop built-in webcam

2.4.2 Image Segmentation Stage

Image segmentation stage is a process of partitioning an image into meaningful regions called segments. This segment will be the region of interest. Image segmentation techniques can be broadly classified into five distinct techniques namely edge detection techniques, thresholding techniques, region-based techniques, clustering-based segmentation techniques, and artificial neural network-based segmentation technique [14]. Aside from the artificial neural network-based segmentation technique, most of this image segmentation stage technique requires human intervention. Human intervention is easily noticed with the usage of LabelImg a special graphical annotation tool [21].

Instead of using the discussed artificial neural network-based segmentation technique or human manual segmentation technique, MediaPipe provide holistic segmentation technique that based on the integration of pose detection, face detection and palm detection [18]–[20]. MediaPipe holistic segmentation technique uses customized MobileNetV2 inspired by MobileNetV1 which was developed based on Convolutional Neural Network (CNN) [22], [23]. With MediaPipe holistic segmentation technique, we managed to retrieve segmentation of human body from Figure 2 as shown in Figure 2.

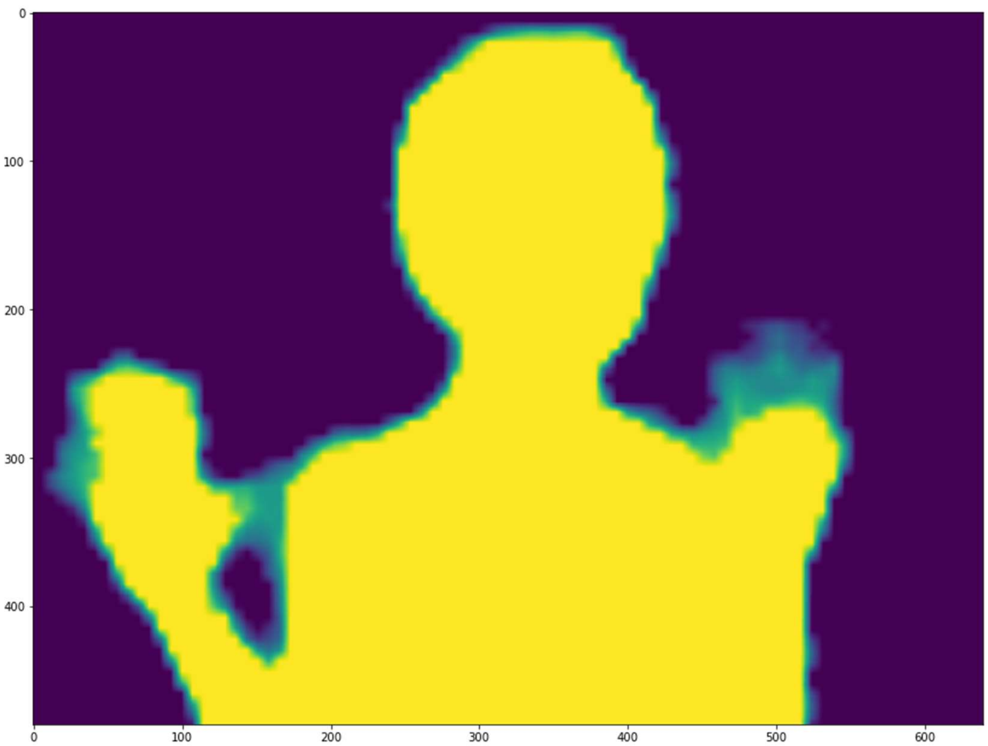

Figure 2. Result of image segmentation with MediaPipe

2.4.3 Image Extraction Stage

Feature extraction uses technique to obtain the highly significant features from acquire input image. The goal is to find the distinctive features in the acquire input image. The compact feature vector represents the interesting parts of an image is extracted with dimensional reduction by removing an irrelevant part to increase learning accuracy and enhance the result’s visibility.

Feature extraction techniques include principal component analysis (PCA), Fourier descriptor (FD), histogram of oriented gradient (HOG), shift-invariant feature transform (SIFT), and speed up robust feature (SURF) [14]. Instead of using the discussed feature extraction techniques, MediaPipe utilizes two stage deep neural network detector-tracker pipeline [22], [23]. MediaPipe feature extraction technique will generate a total of 543 keypoint landmarks as shown in Figure 3.

2.5 Image Classification Techniques

Upon completion of the Image Processing Techniques which include the Image Pre-Processing Stage, Image Segmentation Stage and Image Extraction Stage, predictor algorithm needs to be applied to provides significant meaning to the extracted features.

Machine learning is a subfield of artificial intelligent [1]. Predictor algorithm from machine learning include supervised, semi-supervised, unsupervised and reinforcement [12]. K-Nearest Neighbour (KNN) is the most widely used supervised machine learning as predictor algorithms [24].

Deep learning uses Artificial Neural Network (ANN) such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) [8]. While CNN is sufficient for single image predictor algorithms [4], [25]–[31], RNN provides better accuracy for continuous image predictor algorithms [5], [8], [32]–[34].

Memory usage is the most important feature of RNN that make it possible for better continuous image prediction compared to CNN. However, this memory usage is still insufficient to recall a very long-term data. To resolve this drawback, the long short-term memory (LSTM) model, which provide both long-term memory and short-term memory was created [8], [34]. With LSTM, the ANN can recall both current data and past data [8].

As a conclusion, this paper will use LTSM as predictor algorithm to ensure deep learning can be achieved for continuous sign language image prediction in the proposed Framework for Malaysian Sign Language Recognition using Deep Learning.

Figure 3. Result of feature extraction using keypoint on landmarks with MediaPipe

2.6 Comparison between previous works related to proposed framework initiatives

Based on the discussion on this section, it is facile to perceive the problem in all the previous works discussed. Table 1 shows the problems emerge from all the previous works discussed.

Table 1. Comparison between previous works related to proposed framework.

Reference	Method	Problem	Research gap solution with MediaPipe
[4]	Research built wearable	Not only it is costly to build, but it is also incomplete by itself. The wearable also falls short in capturing non-manual signal.	MediaPipe can provide up to 468 keypoint for face detection, 21 keypoint for right hand detection, 21 keypoint for left hand detection and 33 keypoint for pose detection.
[6]	Research built data glove	Not only it is costly to build, but it is also incomplete by itself. The data glove also falls short in capturing non-manual signal.	
[17]	Kinect camera	It does not come as a standard in any computer system but needed to be bought separately.	MediaPipe is an open source and free framework made available by Google and compatible with any simple webcam.
[15]	Image processing techniques and image classification techniques	It shows poor centroid estimation, poor skin segmentation performance and feature vector issues during image extraction stage.	MediaPipe provide holistic segmentation technique and feature extraction technique.

Therefore, in order to provide a solution for problems emerge from all the previous works discussed it is pertinent to introduce MediaPipe framework in this MSL recognition framework initiative.

3. Methodology

The main objective of this paper is to propose a Framework for Malaysian Sign Language Recognition using Deep Learning. In order to achieve this objective, a framework consisting of three main modules namely learning module, training module and detection module as shown in Figure 4 is proposed.

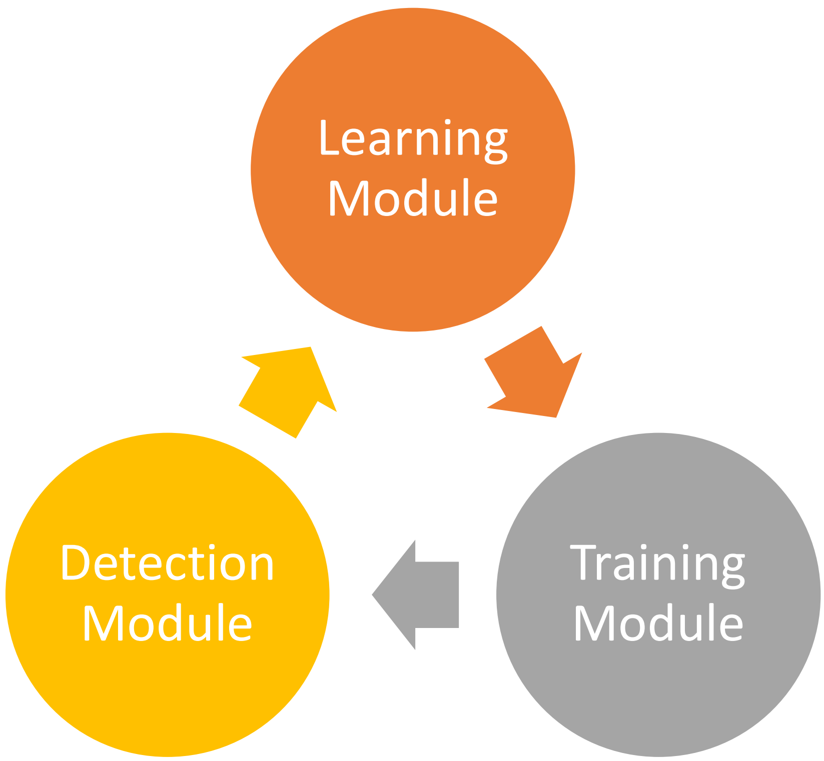

Figure 4. The Framework Main Module for MSL Recognition using Deep Learning

Figure 5 shows the conceptual architecture of the Framework for Malaysian Sign Language Recognition using Deep Learning. As an interpreter, user may choose the learning module. Or as a signer, user may choose the detection module. The training module can only be access by the system if the learning module is completed.

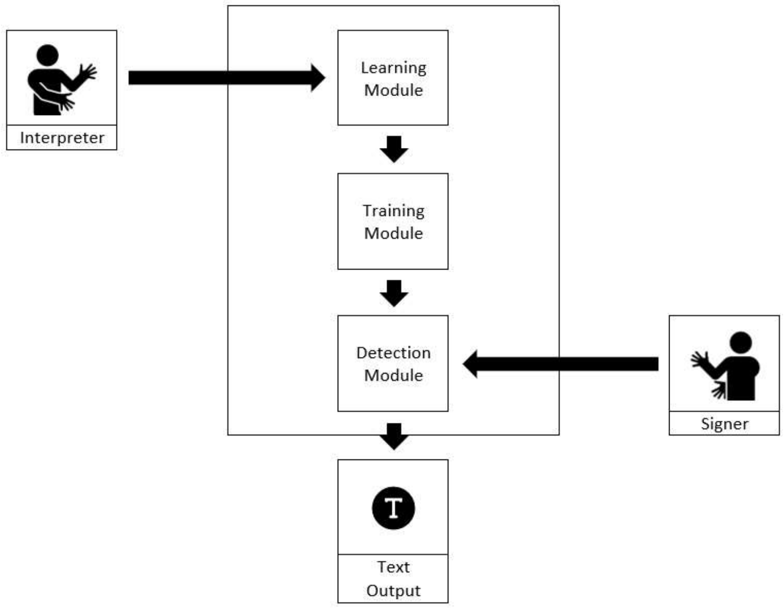

Figure 5. Conceptual Architecture of the Framework

By considering the three main modules namely learning module, training module and detection module and also two type of user which is the interpreter and signer, we derived a complete logical architecture for the proposed Framework for Malaysian Sign Language Recognition using Deep Learning as shown in Figure 6.

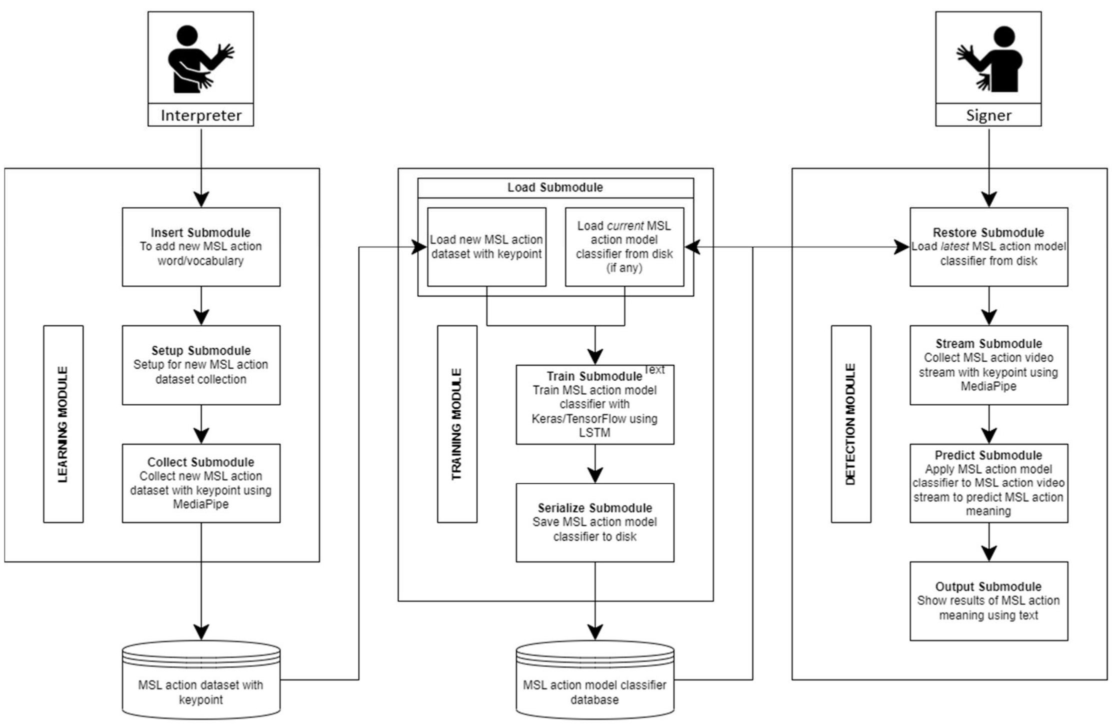

Figure 6. Complete Logical Architecture of the Framework

3.1 Learning Module

As an interpreter, learning module will allow the user to add new vocabulary or word. The learning module itself consists of three submodule which is insert submodule, setup submodule and collect submodule. The output of this module later will be stored in MSL action dataset file and folder. Figure 7 shows the logical architecture of the learning module.

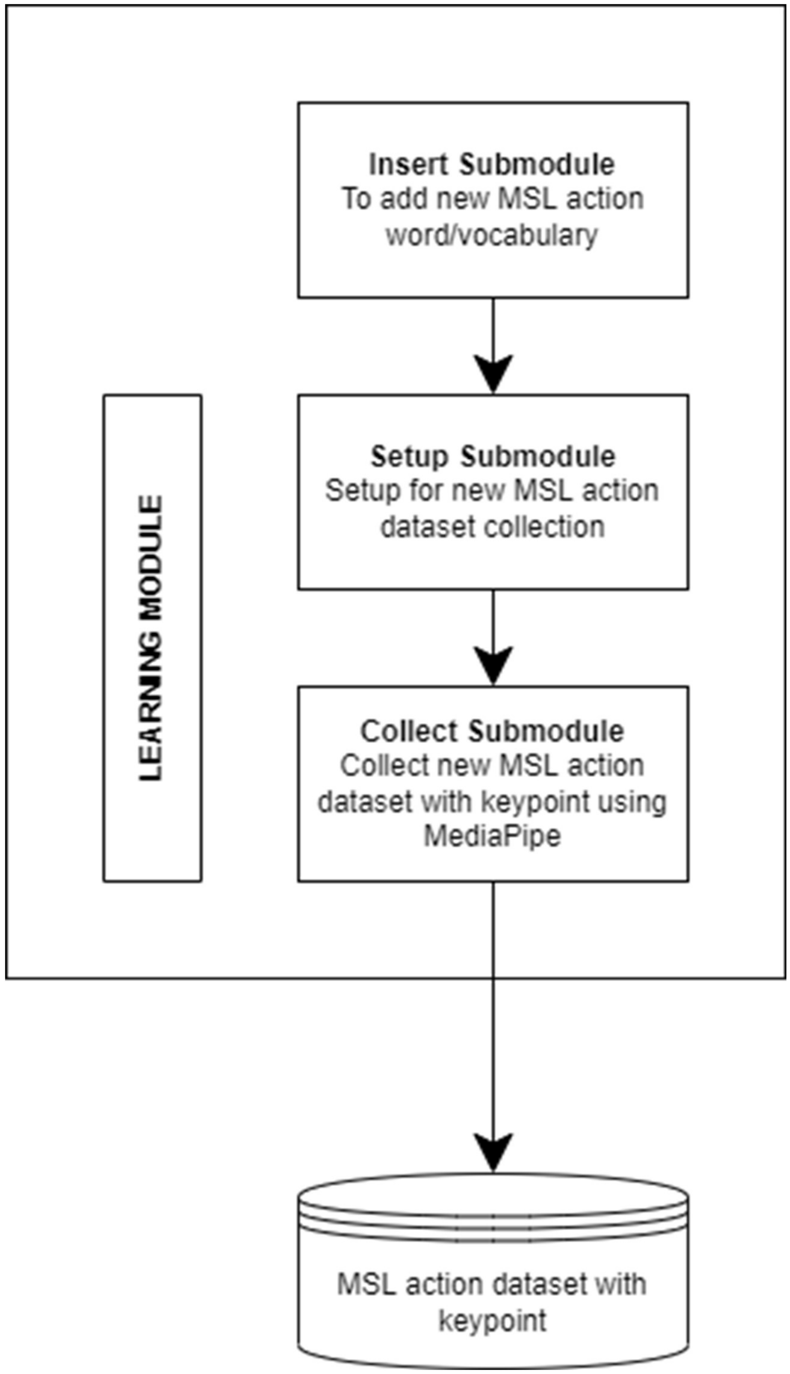

Figure 7. Logical Architecture of The Learning Module

3.1.1 Insert Submodule

Insert submodule will provide an interface for the interpreter to add new MSL action vocabulary or word. To avoid redundancy, insert submodule is also responsible to check if the added MSL action already existed or not in the current MSL action file. If the added MSL action is already existed, a message will be prompt to the interpreter and the added MSL action will be discarded. If the added MSL action does not yet exist in the current MSL action file, the added MSL action will be accepted as new MSL action and the system will proceed with setup submodule.

3.1.2 Setup Submodule

Setup submodule is responsible to create a temporary folder to prepare for the collection of new MSL action dataset. The temporary folder will be name according to the new MSL action. Aside from that, the setup submodule is also responsible to update temporary MSL action file according to the new MSL action. Upon completion of setup submodule, collect module will resume.

3.1.3 Collect Submodule

Collect submodule will provide an interface for the interpreter to capture the new MSL action dataset. This collect submodule is equipped with new feature which make this system different from other existing MSL detection system. This collect submodule will discard the use of costly high-end hardware such as Data Glove, Kinect Camera or Leap Motion Controller.

Additionally, this collect submodule will avoid the usage of laborious manual labelling software tool such as LabelImg. Furthermore, this collect submodule will also omit additional tedious image pre-processing techniques and image segmentation techniques. Instead, this collect submodule will be incorporated with feature extraction techniques during camera function execution. When collect submodule is executed, camera function will be provided with keypoint detection. This keypoint detection will be based on body pose landmark, face landmark, right-hand landmark and left-hand landmark. This is made possible by using MediaPipe.

During execution of collect submodule, 30 set of new MSL action will be collected. Each set will consist of 30 frames of continues new MSL action. Therefore, a complete collection of new MSL action will sum up to 900 frames of images. These frames of images will be stored in MSL action dataset. Upon completion of collect submodule and 30 set times 30 frames of continues new MSL action dataset, training module will be executed.

3.2 Training Module

The training module will allow the system to train the new MSL action dataset with current MSL action model classifier. The training module itself consists of three submodule which is load submodule, train submodule and serialize submodule. The output of this module later will be stored in MSL action model classifier database. Figure 8 shows the logical architecture of the training module.

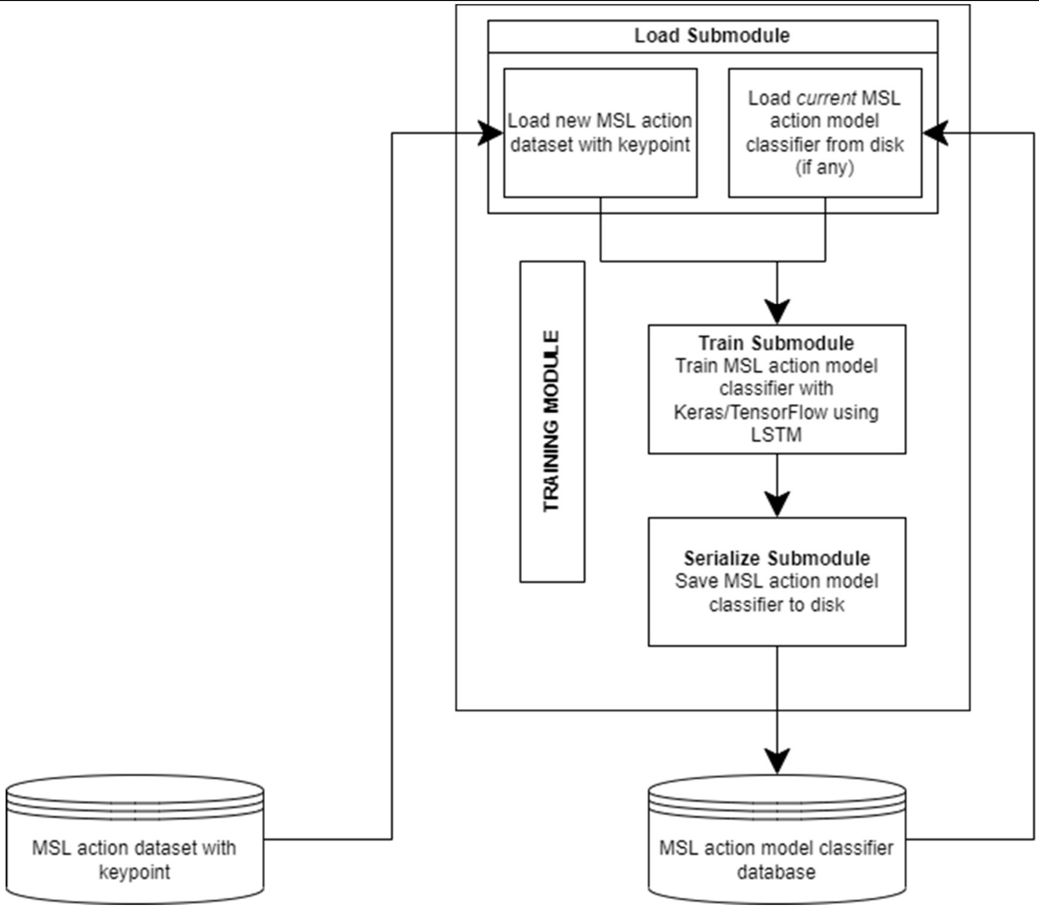

Figure 8. Logical Architecture of The Training Module

3.2.1 Load Submodule

Load submodule will load new MSL action dataset with keypoint. Concurrently, current MSL action model classifier will also be loaded.

3.2.2 Train Submodule

Train submodule is responsible to train current MSL action model classifier with new MSL action dataset. The train submodule is built with LSTM Neural Network model. Current MSL action model classifier with new MSL action from loaded MSL action dataset will be train with LSTM Neural Network model to produce the latest MSL action model classifier.

3.2.3 Serialize Submodule

Serialize submodule will store latest MSL action model classifier in MSL action model classifier database in Hierarchical Data Format version 5 (HDF5), a h5 file. The previous MSL action model classifier h5 file will be archive.

3.3 Detection Module

Detection module will allow a signer to run detection in real time. User may skip the learning module and the training module to proceed with the detection module instantaneously. The detection module itself consists of four submodule which is restore submodule, stream submodule, predict submodule and output submodule. The output of this module will be in text. Figure 9 shows the logical architecture of the detection module.

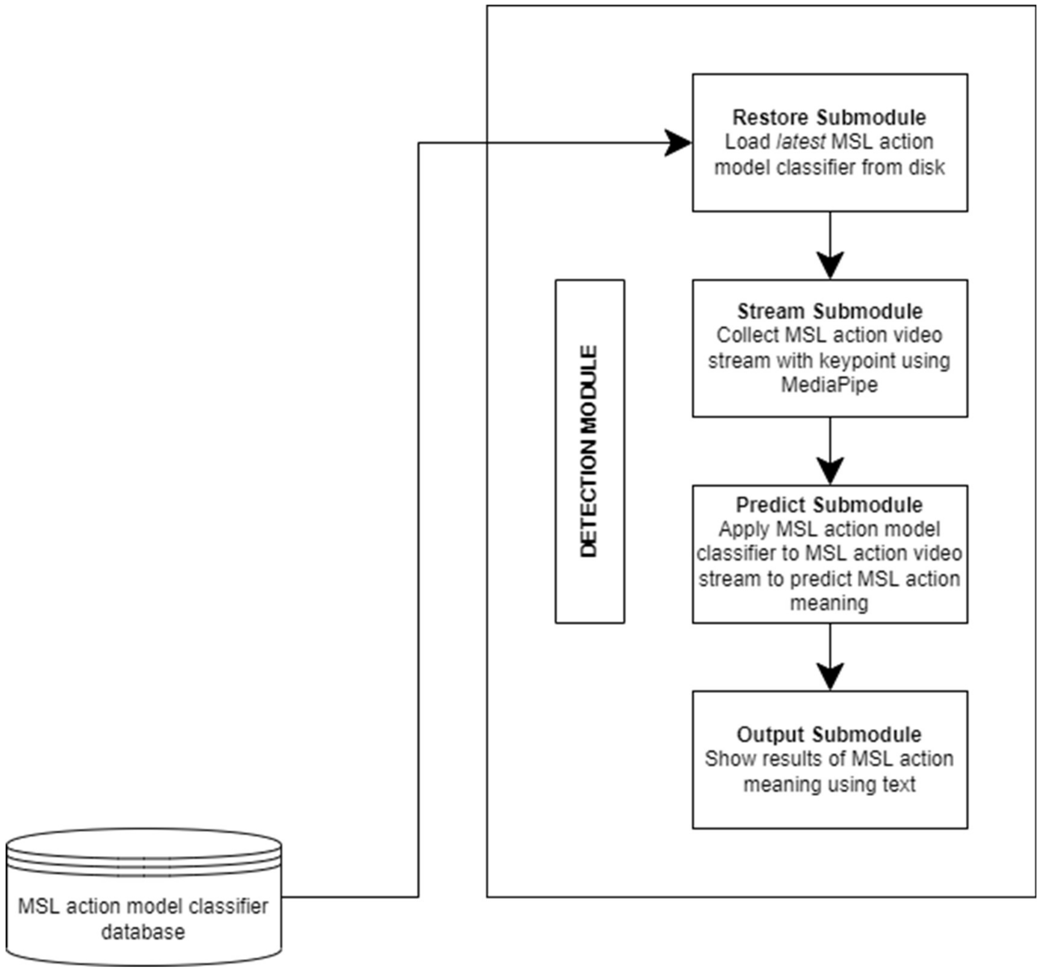

Figure 9. Logical Architecture of The Detection Module

3.3.1 Restore Submodule

Restore submodule will reload latest MSL action model classifier from MSL action model classifier database. The reloaded MSL action model classifier will be used during execution of predict submodule.

3.3.2 Stream Submodule

Stream submodule will provide an interface to capture real time MSL action of a signer. Stream module will also be incorporated with feature extraction techniques during camera function execution. When stream module is executed, camera function will also be provided with keypoint detection. This keypoint detection will also be based on body pose landmark, face landmark, right-hand landmark and left-hand landmark. This is also made possible by using MediaPipe as used in collect submodule in learning module.

3.3.3 Predict Submodule

Predict submodule will extract the captured signer real time MSL action with keypoint from stream submodule and predict the output against reloaded latest MSL action model classifier by restore submodule.

3.3.4 Output Submodule

Based on the result from predict submodule, the captured signer real time MSL action will be translated into text based on the label from MSL action dataset file. Output submodule will display the text output on screen.

4.0 Conclusion

In this paper, a framework for Malaysian Sign Language Recognition using Deep Learning is proposed. In order to achieve this aim, the usage of MediaPipe framework was introduced. MediaPipe framework improved image acquisition and simplified image processing stage tremendously. Long short-term memory (LSTM) artificial neural network (ANN) is used as training algorithm in training module and prediction algorithm in detection module. With this framework initiative, it is anticipated that communication barrier between deaf and hard of hearing people who use MSL with person without disabilities who do not use MSL can be overcome.

Acknowledgements

The authors gratefully acknowledge the Universiti Teknologi MARA (UiTM), Perak branch.

Conflict of Interest

The authors declare no conflict of interest in the subject matter or materials discussed in this manuscript.

References

[1] E. Kavlakoglu, “AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What’s the Difference?,” 2020. [Online]. Available: https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks.

[2] M. R. Islam, S. Barua, M. U. Ahmed, S. Begum, and G. Di Flumeri, “Deep Learning for Automatic EEG Feature Extraction: An Application in Drivers’ Mental Workload Classification,” *Commun. Comput. Inf. Sci.*, vol. 1107, no. November, pp. 121–135, 2019.

[3] T. Muralidharan and N. Nissim, “Improving malicious email detection through novel designated deep-learning architectures utilizing entire email,” *Neural Networks*, 2022.

[4] K. Van Murugiah, G. Subhashini, and R. Abdulla, “Wearable IoT based Malaysian sign language recognition and text translation system,” *J. Appl. Technol. Innov.*, vol. 5, no. 4, pp. 51–58, 2021.

[5] S. M. Kamal, Y. Chen, S. Li, X. Shi, and J. Zheng, “Technical Approaches to Chinese Sign Language Processing: A Review,” *IEEE Access*, vol. 7, pp. 96926–96935, 2019.

[6] A. Z. Shukor, M. F. Miskon, M. H. Jamaluddin, F. Bin Ali Ibrahim, M. F. Asyraf, and M. B. Bin Bahar, “A New Data Glove Approach for Malaysian Sign Language Detection,” *Procedia Comput. Sci.*, vol. 76, no. Iris, pp. 60–67, 2015.

[7] Sutarman, M. B. A. Majid, J. B. M. Zain, and A. Hermawan, “Recognition of Malaysian Sign Language using skeleton data with Neural Network,” *Proc. - 2015 Int. Conf. Sci. Inf. Technol. Big Data Spectr. Futur. Inf. Econ. ICSITech 2015*, no. October, pp. 231–236, 2016.

[8] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, “Deepsign: Sign Language Detection and Recognition Using Deep Learning,” *Electronics*, vol. 11, no. 11, p. 1780, 2022.

[9] A. H. Alrubayi *et al.*, “A pattern recognition model for static gestures in malaysian sign language based on machine learning techniques,” *Comput. Electr. Eng.*, vol. 95, no. April 2020, p. 107383, 2021.

[10] T. S. Tan, A. K. Ariff, S. H. Salleh, K. S. Siew, and S. H. Leong, “Wireless data gloves Malay sign language recognition system,” *2007 6th Int. Conf. Information, Commun. Signal Process. ICICS*, pp. 3–6, 2007.

[11] N. M. Kakoty and M. D. Sharma, “Recognition of Sign Language Alphabets and Numbers based on Hand Kinematics using A Data Glove,” *Procedia Comput. Sci.*, vol. 133, pp. 55–62, 2018.

[12] M. S. Amin, S. T. H. Rizvi, and M. M. Hossain, “A Comparative Review on Applications of Different Sensors for Sign Language Recognition,” *J. Imaging*, vol. 8, no. 4, pp. 1–48, 2022.

[13] D. Kumar Choudhary, R. Singh, and D. Kamathania, “Sign Language Recognition System,” *4TH Int. Conf. Innov. Comput. Commun. (ICICC 2021)*, no. ICICC, pp. 4–7, 2021.

[14] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, “Machine learning methods for sign language recognition: A critical review and analysis,” *Intell. Syst. with Appl.*, vol. 12, pp. 1–36, 2021.

[15] F. Wong, G. Sainarayanan, W. M. Abdullah, A. Chekima, F. E. Jupirin, and Y. F. A. Gaus, “Software-based Malaysian sign language recognition,” *Adv. Intell. Syst. Comput.*, vol. 182 AISC, pp. 297–306, 2013.

[16] M. Karabasi, Z. Bhatti, and A. Shah, “A model for real-time recognition and textual representation of malaysian sign language through image processing,” *Proc. - 2013 Int. Conf. Adv. Comput. Sci. Appl. Technol. ACSAT 2013*, no. December, pp. 195–200, 2013.

[17] M. Karbasi, A. Zabidi, I. M. Yassin, A. Waqas, and Z. Bhatti, “Malaysian sign language dataset for automatic sign language recognition system,” *J. Fundam. Appl. Sci.*, vol. 9, no. 4S, p. 459, 2018.

[18] C. Lugaresi *et al.*, “MediaPipe: A Framework for Building Perception Pipelines,” *Google Res.*, pp. 1–9, 2019.

[19] C. Lugaresi *et al.*, “MediaPipe: A Framework for Perceiving and Processing Reality,” *Google Res.*, pp. 1–4, 2019.

[20] F. Zhang *et al.*, “MediaPipe Hands: On-device Real-time Hand Tracking,” *Google Res.*, pp. 1–5, 2020.

[21] N. A. Mohammad, S. T. Kian, F. S. Chin, H. S. Toong, and R. Abdul Rahim, “Development of a Malaysian Sign Language interpreter using image recognition for the community to understand the deaf,” *Elektr. J. Electr. Eng.*, vol. 20, no. 2-3 SE-Articles, pp. 70–72, 2021.

[22] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, “BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs,” pp. 3–6, 2019.

[23] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “BlazePose: On-device Real-time Body Pose tracking,” 2020.

[24] S. Shekhar, N. Hoque, and D. K. Bhattacharyya, “PKNN-MIFS: A Parallel KNN Classifier over an Optimal Subset of Features,” *Intell. Syst. with Appl.*, vol. 14, p. 200073, 2022.

[25] D. K. Singh, “3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling,” *Procedia CIRP*, vol. 189, pp. 76–83, 2021.

[26] S. Katoch, V. Singh, and U. S. Tiwary, “Indian Sign Language recognition system using SURF with SVM and CNN,” *Array*, vol. 14, no. March, p. 100141, 2022.

[27] A. Liew, S. Lien, and L. K. Yin, “Gesture Recognition-Malaysian Sign Language Recognition using Convolutional Neural Network,” in *International Conference on Digital Transformation and Applications (ICDXA) 2020*, 2020, pp. 1–6.

[28] M. A. M. M. Asri, Z. Ahmad, I. A. Mohtar, and S. Ibrahim, “A real time Malaysian sign language detection algorithm based on YOLOv3,” *Int. J. Recent Technol. Eng.*, vol. 8, no. 2 Special Issue 11, pp. 651–656, 2019.

[29] K. Wangchuk, P. Riyamongkol, and R. Waranusast, “Real-time Bhutanese Sign Language digits recognition system using Convolutional Neural Network,” *ICT Express*, vol. 7, no. 2, pp. 215–220, 2021.

[30] Suharjito, N. Thiracitta, and H. Gunawan, “SIBI Sign Language Recognition Using Convolutional Neural Network Combined with Transfer Learning and non-trainable Parameters,” *Procedia Comput. Sci.*, vol. 179, no. 2019, pp. 72–80, 2021.

[31] A. Ardiansyah, B. Hitoyoshi, M. Halim, N. Hanafiah, and A. Wibisurya, “Systematic Literature Review: American Sign Language Translator,” *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 541–549, 2021.

[32] S. Subburaj and S. Murugavalli, “Measurement : Sensors Survey on sign language recognition in context of vision-based and deep learning,” *Meas. Sensors*, vol. 23, no. June, p. 100385, 2022.

[33] S. Alashhab, A. J. Gallego, and M. Á. Lozano, “Efficient Gesture Recognition for the Assistance of Visually Impaired People using Multi-Head Neural Networks,” *Eng. Appl. Artif. Intell.*, vol. 114, no. June, p. 105188, 2022.

[34] Ü. Atila and F. Sabaz, “Turkish lip-reading using Bi-LSTM and deep learning models,” *Eng. Sci. Technol. , an Int. J.*, no. xxxx, 2022.

Biography of author

Picture	Biography	Authorship contribution
	Imran Md Jelas is a lecturer in Universiti Teknologi MARA (UiTM), Perak branch, Tapah campus. He is currently pursuing his PhD degree in Universiti Kebangsaan Malaysia. His current research interests include big data and deep learning.	Design the research work and drafting article and camera-ready.