SEMARAK ILMU PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Malaysian Sign Language Detection with Convolutional Neural Network for Disabilities

Iffat Adib Kurnia Mohd Ramzi[1], Mohd Syahril Ramadhan Mohd Saufi[1,*], Mohd Zarhamdy Md. Zain[1], Mastura Ab Wahid[1], Mat Hussin Ab Talib[1], Nur Fathiah Waziralilah[2], Kamarul Arifin Hassan[1]

1 Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
2 Faculty of Mechanical Engineering, Alfaisal University, Riyadh 11533, Saudi Arabia

ABSTRACT

Sign language is the main form of communication used by deaf people. Most of their activities like speaking and learning involved sign language. In Malaysia, instead of a deaf person, only a few people know sign language because it's not part of the co-curriculum. Hence, normal people need to know sign language so that they can communicate with deaf people effectively. However, learning sign language requires a lot of effort and takes years of study. Therefore, this research tends to initiate the sign language image detection system that utilizes the deep learning model. The initial effort involves the development of an image detection system that can identify the alphabet of sign language from A to Z. There will be 26 alphabets that contribute to 26 classes of image detection using the Convolutional Neural Network. 85 images will be collected for each alphabet and the total data will be distributed into 80% and 20% for training and testing the CNN model, respectively. The result showed that the proposed system produced more than 95% detection accuracy of the alphabet based on finger sign gestures. For future works, the proposed system will be evaluated with a more complex dataset in terms of words.

Keywords:
Malaysia sign language; Convolutional Neural Network; web-based application; TensorFlow

1. Introduction

Sign Language (SL) is the only way of communication for deaf people when verbal communication is impossible for them. SL employs visuals to communicate through body gestures, particularly of the hands and arms [1]. Sign language will be used to communicate with or among handicapped people since they have no other means of communication [2,3]. To become fluent in sign language normally takes several years of consistent study and practice. Only a small percentage of Malaysians can utilize sign language because it was not included in the standard curriculum. Sign language is essential for parents of deaf children who need to learn to converse with their children. According to the former Health Minister, there were 40,743 registered hearing-impaired people in Malaysia as of December

* *Corresponding author.*
E-mail address: syahrilramadhan@mail.fkm.utm.my

https://doi.org/10.37934/araset.58.1.294302

2021 [4]. Hence, providing an effective communication platform for sign language is important to promote better communication between normal and disabled persons.

Learning sign language may help to increase the brain's talents and thinking, as well as memorizing capacity, creativity, and communication skills. The benefits of learning sign language are similar to those of learning a second language. Learning any language, whether signed or spoken, is extremely beneficial to the brain. The world today has several hundred different varieties of sign language; Malaysia also has its sign language [5]. Malaysian Sign Language (BIM) is used by the majority of the deaf people in this nation, as illustrated in Figure 1. It is an official language that the government has approved for usage in community and education. The government does not leave the deaf people by broadcasting a BIM interpreter in sign language.

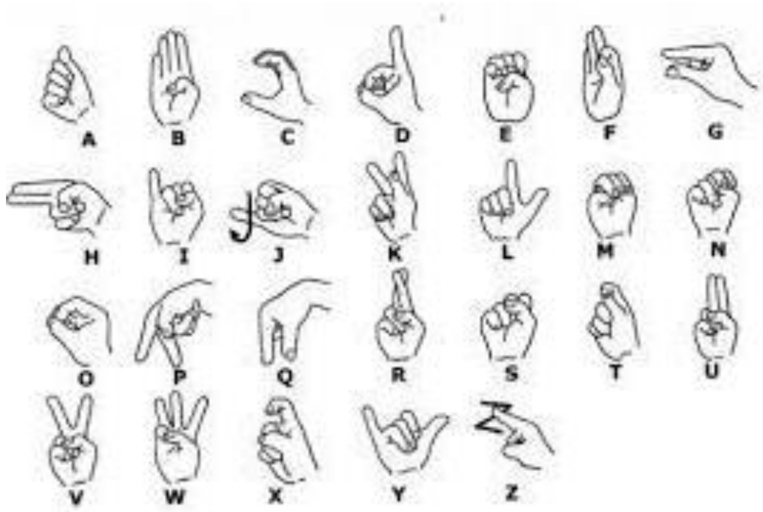

Fig. 1. Malaysian sign language (BIM)

Much research has been conducted in recent years to solve the challenge of sign language recognition. However, there is no perfect answer to the problem in today's society. This challenge has been handled in two ways: through contact-based systems such as sensor gloves or through vision-based systems that simply employ cameras [6,7]. As a result, the introduction of deep learning, which is significantly less expensive and more user-friendly, influences other researcher specialists to believe that the problem of sign language identification will be entirely addressed [8-11]. Deep learning is an algorithm that mimics how people learn. Sign language recognition may be accomplished by utilizing deep learning methods in picture classification, object identification, and natural language processing [11-16].

Convolutional Neural Networks (CNN) is the common deep learning model that has been used in image detection analysis and has been proven to be effective in object detection. Based on García-Aguilar *et al.,* [17] the CNN model can be used in online and real-time analysis which is an important element in developing real-time detection of sign language [17-21]. Hence, CNN will be used as a prediction model for sign language images.

This project focuses on image detection of sign language to help normal people understand sign language by using the system of deep learning embedded in web-based. The scope of this project is detecting the sign language alphabet A to Z only, collecting images from surrounding people, using Python and TensorFlow platforms, and finally implementing the model web-based. The rest of the paper is organized as follows:

i. Section 2 gives a brief overview of the Convolutional Neural Network.
ii. Section 3 provides a discussion of data collection and data preparation.
iii. Section 4 discusses the proposed methodology of alphabet sign detection based on the CNN model.
iv. Section 5 evaluates the proposed framework and analyses experiment results.
v. Section 6 concludes the paper.

2. The Proposed Model with Convolutional Neural Network (CNN)

The proposed model required the development of a CNN model with TensorFlow, and dataset acquisition to collect real-world data and then implement it on the web. This project aims to construct a Convolutional Neural Network (CNN) model with above 90% accuracy using TensorFlow API for web-based use. The training and testing procedure of the CNN model is shown in Figure 2 and this study is limited to the alphabet sign language.

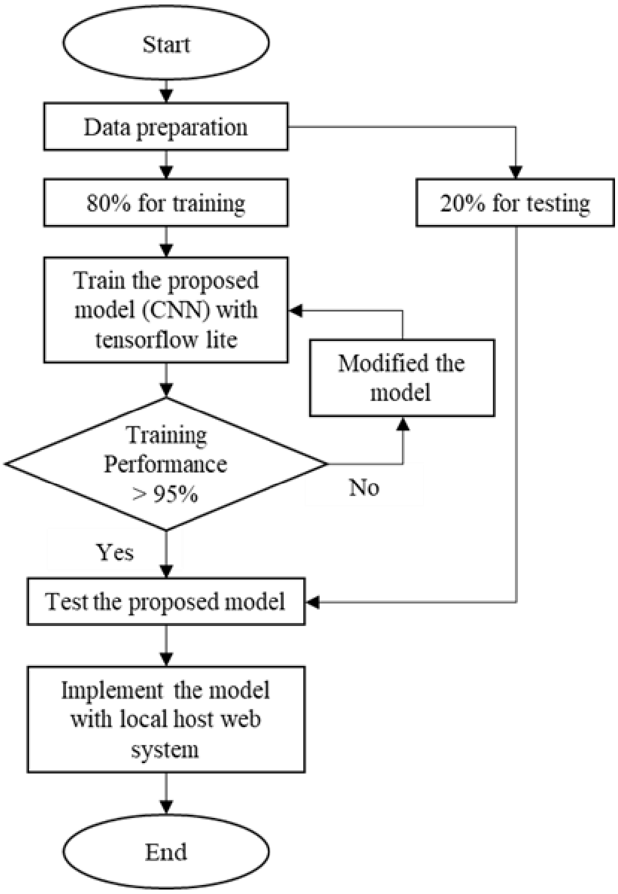

Fig. 2. The training and testing procedure of the CNN model

2.1 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a kind of deep neural network that has been mainly designed for 2D images. Due to its convolutional function, it is superior in image detection compared to the artificial neural network. The convolutional layer is a specific type of layer that can extract complex features from the input data.

During the training phase, the CNN network will learn the pattern between the input and the output using the backpropagation algorithm. The prediction data will then be compared with the label data in which the difference between the label and the predicted value is called the prediction error as defined in Eq. (3). The training aims to reduce this error by reducing the prediction value. According to Eq. (1) and Eq. (2), the prediction value yt is affected by parameter values called weight and bias. Therefore, these parameters will be tuned to reduce the prediction value yt. Table 1 represents the size of the input image and the hyperparameter of the CNN model. In this research, the image size and weight regularizer will be changed to study the behaviour of the CNN model toward sign language images.

$$h_t = f(wx + b_h) \tag{1}$$

$$y_t = f\left(wh_t + b_y\right) \tag{2}$$

$$E_t = \left(d_t - y_t\right)^2 \tag{3}$$

Table 1
The size of the input image and the hyperparameter of the CNN model

	First analysis		Second analysis	
Height x width of the image	320x320	640x640	320x320	
Weight regularizer	1×10^{-4}		1×10^{-4}	1×10^{-7}
Minimum depth	16			
Batch size	4			
Learning rate	0.08			
Momentum	0.9			

3. Data Collection and Data Preparation

The data was collected with self-captured of the own images. For the self-captured images, the background of the data is varied to ensure the diversity of the datasets to improve detection accuracy. 85 images were collected from self-captured and online for each alphabet. The total dataset for all alphabet signs is 2,210 with varying backgrounds, angles, lighting, and individuals. Figure 3 shows the whole dataset acquired with self-captured images. In developing the CNN models, the dataset was divided into an 80% training set and a 20% testing set.

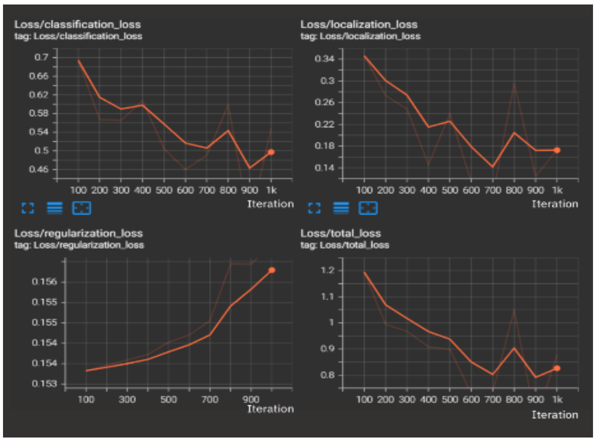

Fig. 3. Training result of the CNN model on 320x320 and 640x640 images

4. Result and Discussion

The 2,210-image dataset was split into 80% and 20% for training and testing. During image processing, a 640x480 image was captured. The images were resized and it was transformed into a greyscale image to reduce the CNN's model computation time. It is known that working with RGB images is computationally intensive. According to Figure 4, the detection did not require a colour image since the important features of the sign language can be examined via the line and shape of the hand gesture. The training is complete and tested on the desktop before being implemented on the web. The precision and loss will be used to measure the performance of the CNN model. Based on the precision and loss graph, the model behaviour can be determined as either overfitting, good fit, or underfitting. If the graph displays poor performance, an adjustment needs to be made in terms of hyperparameter tuning and the size of the dataset. The common hyperparameter that needs to be tuned is the weight regularizer which this hyperparameter will help to control the weight connection in the CNN network. During the CNN model development, the image position, angle, lightning, and density of the dataset to generalize training and avoid overfitting. There are two analyses involved where the image size and weight regularizer were changed.

Model A (Detection/Precision)	Model B (Detection/Precision)

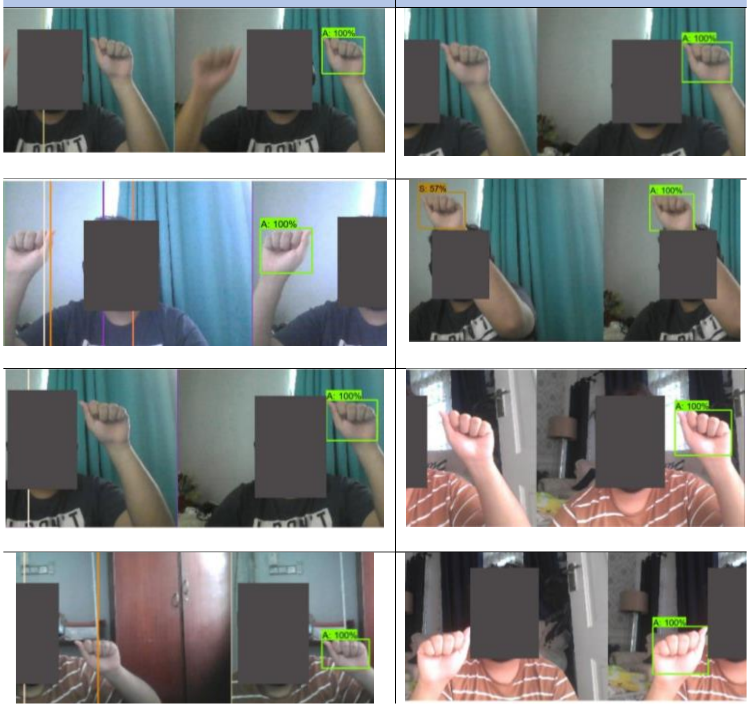

Fig. 4. Comparison of testing performance between model A and model B

4.1 CNN Results with Different Weight Regularizer Value

For the second analysis, the model was trained with 320x320 images, and the weight regularization in the CNN model was changed to 1x10-7. Results from Figure 5 proved that the performance of this model is better than the previous model (Figure 3) since the loss regularization shows a decrement trend. The smaller weight regularization between nodes enables the model to avoid overfitting.

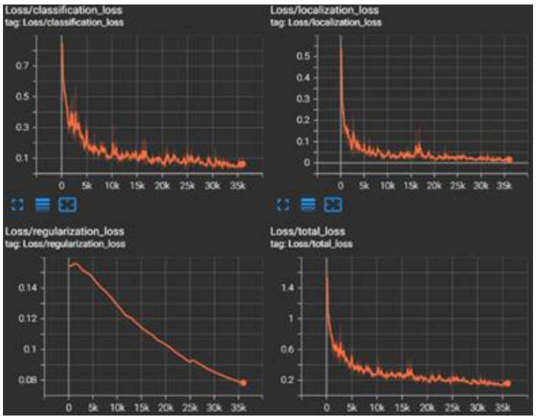

Fig. 5. Training result of the CNN model on 320x320 and 640x640 images

Hence, the CNN model with lower weight regularization is chosen as the final model since it performed well in detecting all the sign language alphabets which achieved this project objective. Even though the model was trained with greyscale images, it still can be tested with RGB images and produced an accurate prediction as shown in Figure 6.

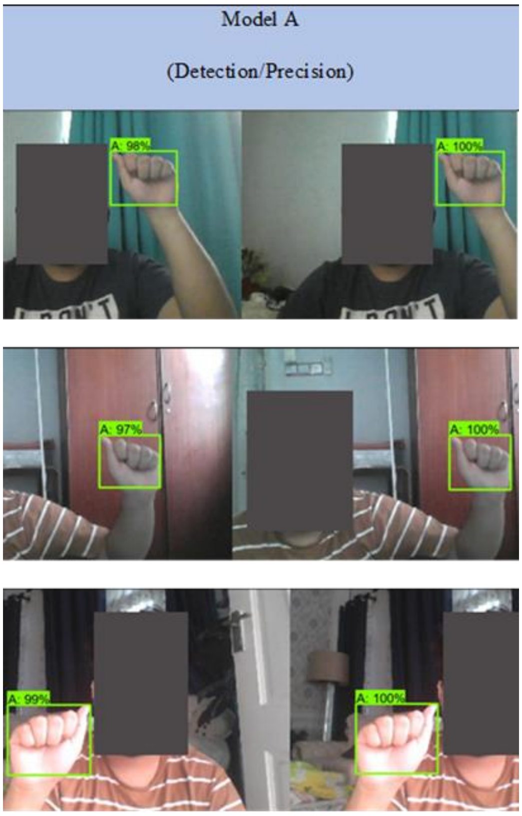

Fig. 6. Testing performance of model A with 1x10^{-4} weight regularizer value

4.2 Implementation in Local Web Host System

The CNN model with tuned weight regularizer will be converted to JavaScript for Web-based application. JavaScript was utilized to construct a dynamic, interactive model for IBM Cloud. IBM Cloud is a portable application and data platform for AI, hybrid multi-cloud platforms, and advanced data. The trained CNN model from the TensorFlow model was transferred to TensorFlow.js. TensorFlow.js. TensorFlow.js is a JavaScript machine-learning library that makes machine-learning models web-browser-friendly. The conversion worked in the browser is shown in Figure 7 and Figure 8. It correctly identified all sign language alphabet classes. Above 80 to 90% of classes were accurately detected. However, after converting to a web-based application the detection speed is slower.

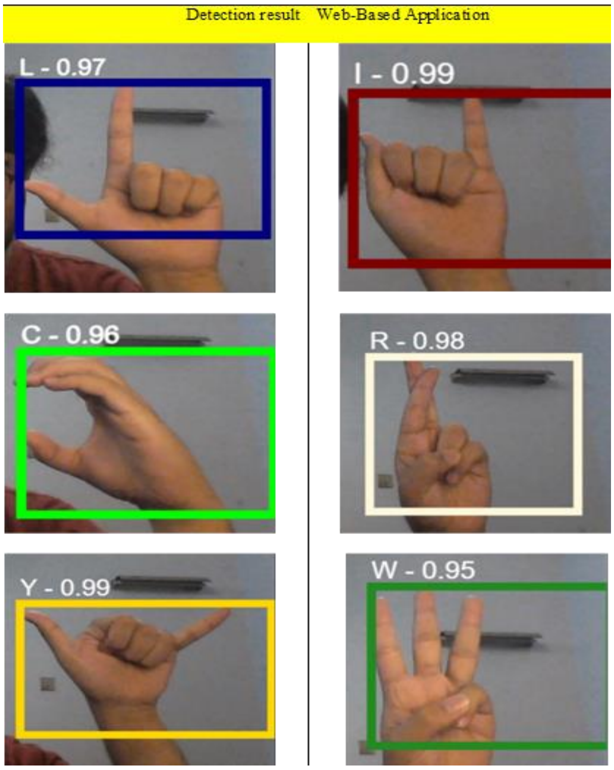

Fig. 7. Detection result in web-based application

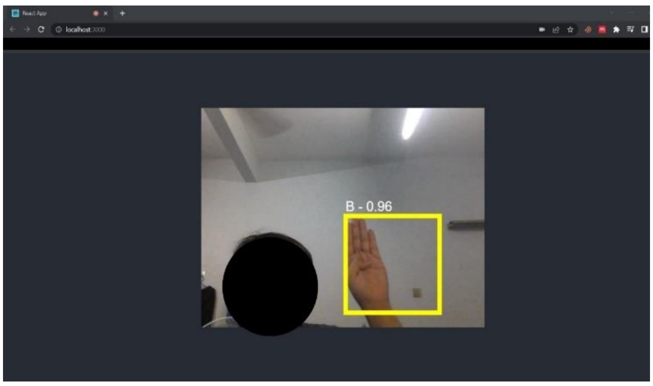

Fig. 8. Web-based application

5. Conclusion

The project proposed the detection of Malaysian sign language using the CNN model and transforming it into a web-based application. This research concludes that the CNN model's overfitting, underfitting, or optimal status may be determined by its loss function graph. Adjusting the dataset size, number of training iterations, weight regularisation, and input image size can affect the CNN model's performance. The weight regularisation tuning can improve the model's recall and precision. The proposed model produced a good detection of the alphabet sign. The project will be further conducted towards the words of the sign language instead of the alphabet and the system will be developed as a smartphone application.

Acknowledgement
This work was financially supported by the Universiti Teknologi Malaysia under the Universiti Teknologi Malaysia Fundamental Research grant with scheme number Q.J130000.3851.22H06.

References
[1] Maarif, Haris Al Qodri, Rini Akmeliawati, and Sara Bilal. "Malaysian Sign Language database for research." In *2012 International Conference on Computer and Communication Engineering (ICCCE)*, pp. 798-801. IEEE, 2012. https://doi.org/10.1109/ICCCE.2012.6271327
[2] Saiful Bahri, Iffah Zulaikha, Sharifah Saon, Abd Kadir Mahamad, Khalid Isa, Umi Fadlilah, Mohd Anuaruddin Bin Ahmadon, and Shingo Yamaguchi. "Interpretation of Bahasa Isyarat Malaysia (BIM) Using SSD-MobileNet-V2 FPNLite and COCO mAP." *Information* 14, no. 6 (2023): 319. https://doi.org/10.3390/info14060319
[3] Mahmoud, Nourelhoda M., Hassan Fouad, and Ahmed M. Soliman. "Smart healthcare solutions using the internet of medical things for hand gesture recognition system." *Complex & intelligent systems* 7 (2021): 1253-1264. https://doi.org/10.1007/s40747-020-00194-9
[4] Muthiah, W. "KJ: Over 40,000 hearing-impaired people registered in Malaysia." *The Star*, (2022). https://www.thestar.com.my/news/nation/2022/03/03/kj-over-40000-hearing-impaired-people-registered-in-malaysia
[5] Alrubayi, Ali H., Mohamed Aktham Ahmed, A. A. Zaidan, Ahmed Shihab Albahri, B. B. Zaidan, Osamah Shihab Albahri, Abdullah Hussein Alamoodi, and Mamoun Alazab. "A pattern recognition model for static gestures in malaysian sign language based on machine learning techniques." *Computers and Electrical Engineering* 95 (2021): 107383. https://doi.org/10.1016/j.compeleceng.2021.107383
[6] Pathak, Ajeet Ram, Manjusha Pandey, and Siddharth Rautaray. "Application of deep learning for object detection." *Procedia computer science* 132 (2018): 1706-1717. https://doi.org/10.1016/j.procs.2018.05.144
[7] Shukor, Ahmad Zaki, Muhammad Fahmi Miskon, Muhammad Herman Jamaluddin, Fariz bin Ali, Mohd Fareed Asyraf, and Mohd Bazli bin Bahar. "A new data glove approach for Malaysian sign language detection." *Procedia Computer Science* 76 (2015): 60-67. https://doi.org/10.1016/j.procs.2015.12.276
[8] Khan, Rehman Ullah, Hizbullah Khattak, Woei Sheng Wong, Hussain AlSalman, Mogeeb AA Mosleh, and Sk Md Mizanur Rahman. "Intelligent Malaysian Sign Language Translation System Using Convolutional-Based Attention Module with Residual Network." *Computational Intelligence and Neuroscience* 2021, no. 1 (2021): 9023010. https://doi.org/10.1155/2021/9023010
[9] Neethu, P. S., R. Suguna, and Divya Sathish. "An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks." *Soft Computing* 24, no. 20 (2020): 15239-15248. https://doi.org/10.1007/s00500-020-04860-5
[10] Mujahid, Abdullah, Mazhar Javed Awan, Awais Yasin, Mazin Abed Mohammed, Robertas Damaševičius, Rytis Maskeliūnas, and Karrar Hameed Abdulkareem. "Real-time hand gesture recognition based on deep learning YOLOv3 model." *Applied Sciences* 11, no. 9 (2021): 4164. https://doi.org/10.3390/app11094164
[11] Sharma, Sakshi, and Sukhwinder Singh. "Vision-based hand gesture recognition using deep learning for the interpretation of sign language." *Expert Systems with Applications* 182 (2021): 115657. https://doi.org/10.1016/j.eswa.2021.115657
[12] Alani, Ali A., and Georgina Cosma. "ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition." *Indonesian journal of electrical engineering and computer science* 22 (2021). https://doi.org/10.11591/ijeecs.v22.i2.pp1096-1107
[13] Oudah, Munir, Ali Al-Naji, and Javaan Chahl. "Hand gesture recognition based on computer vision: a review of techniques." *journal of Imaging* 6, no. 8 (2020): 73. https://doi.org/10.3390/jimaging6080073
[14] Mohamed, Noraini, Mumtaz Begum Mustafa, and Nazean Jomhari. "A review of the hand gesture recognition system: Current progress and future directions." *IEEE access* 9 (2021): 157422-157436. https://doi.org/10.1109/ACCESS.2021.3129650
[15] Tan, Yong Soon, Kian Ming Lim, Connie Tee, Chin Poo Lee, and Cheng Yaw Low. "Convolutional neural network with spatial pyramid pooling for hand gesture recognition." *Neural Computing and Applications* 33 (2021): 5339-5351. https://doi.org/10.1007/s00521-020-05337-0

[16] Lim, Kian Ming, Alan Wee Chiat Tan, Chin Poo Lee, and Shing Chiang Tan. "Isolated sign language recognition using convolutional neural network hand modelling and hand energy image." *Multimedia Tools and Applications* 78 (2019): 19917-19944. https://doi.org/10.1007/s11042-019-7263-7

[17] García-Aguilar, Iván, Jorge García-González, Rafael Marcos Luque-Baena, and Ezequiel López-Rubio. "Automated labeling of training data for improved object detection in traffic videos by fine-tuned deep convolutional neural networks." *Pattern Recognition Letters* 167 (2023): 45-52. https://doi.org/10.1016/j.patrec.2023.01.015

[18] Mohanty, Aparna, Sai Saketh Rambhatla, and Rajiv Ranjan Sahay. "Deep gesture: static hand gesture recognition using CNN." In *Proceedings of International Conference on Computer Vision and Image Processing: CVIP 2016, Volume 2*, pp. 449-461. Springer Singapore, 2017. https://doi.org/10.1007/978-981-10-2107-7_41

[19] Wu, Xiao Yan. "A hand gesture recognition algorithm based on DC-CNN." *Multimedia Tools and Applications* 79, no. 13 (2020): 9193-9205. https://doi.org/10.1007/s11042-019-7193-4

[20] Park, Gyutae, Vasantha Kumar Chandrasegar, and Jinhwan Koh. "Accuracy enhancement of hand gesture recognition using CNN." *IEEE Access* 11 (2023): 26496-26501. https://doi.org/10.1109/ACCESS.2023.3254537

[21] Palaniapan, Sri Ganes, and Ashley Ng Sok Choo. "A Hybrid Model for Human Action Recognition Based on Local Semantic Features." *Journal of Advanced Research in Computing and Applications* 33, no. 1 (2023): 7-21. https://doi.org/10.37934/arca.33.1.721