© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Recognition of Radar-Based Deaf Sign Language Using Convolution Neural Network

M. D. H. Dol Malik[1,4], W. Mansor[1,2,3*], N. E. Abdul Rashid[1,2] and M. Z. U. Rahman[5]

[1]School of Electrical Engineering, College of Engineering,
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA

[2]Microwave Research Institute Engineering Complex,
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA

[3]Computational Intelligence Detection RIG, Health & Wellness ReNeU,
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA

[4]TNB Integrated Learning Solution Sdn Bhd-ILSAS, Jalan Ikram-Uniten,
Institut Latihan Sultan Ahmad Shah, 43650 Bandar Baru Bangi, Selangor, MALAYSIA

[5]Department of Electronics and Communication Engineering,
Koneru Lakshmaiah Education Foundation, Guntur-522502, A.P., INDIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2023.15.03.012
Received 6 November 2022; Accepted 18 April 2023; Available online 15 August 2023

Abstract: The difficulties in the communication between the deaf and normal people through sign language can be overcome by implementing deep learning in the gestures signal recognition. The use of the Convolution Neural Network (CNN) in distinguishing radar-based gesture signals of deaf sign language has not been investigated. This paper describes the recognition of gestures of deaf sign language using radar and CNN. Six gestures of deaf sign language were acquired from normal subjects using a radar system and processed. Short-time Fourier Transform was performed to extract the gestures features and the classification was performed using CNN. The performance of CNN was examined using two types of inputs; segmented and non-segmented spectrograms. The accuracy of recognising the gestures is higher (92.31%) using the non-segmented spectrograms compared to the segmented spectrogram. The radar-based deaf sign language could be recognised accurately using CNN without segmentation.

Keywords: Radar, deep learning, Short-Time Fourier Transform (STFT), gestures, classification

1. Introduction

Sign languages are the viable communication methods in deaf communities, hearing people who are unable to speak and people who have deaf family members. The sign languages could not be understood by most normal people who have not studied or acquired them. Thus, there is a barrier in communication between deaf and normal people where some messages could not be delivered and interpreted correctly. The difficulties in this communication can be overcome by introducing automatic recognition of hand gestures using artificial intelligence [1-6].

The detection of several gestures such as palm-clench, palm-open, hand-backward, hand-forward, pinch-part, pinch together, swipe right, swipe left, zoom in, zoom out, pushing, pulling, lifting a hand, and shaking hand have been

*Corresponding author: wahidah231@uitm.edu.my

investigated using radar [7-10]. Radar is an electromagnetic sensor that works by transmitting electromagnetic energy to a target and, then recording the echoes returned from the target [11]. The capability of radar to detect an object, from as small as rain to a larger object like aircraft [12-14], makes it suitable to be used to recognise gestures. In our previous work, we have analysed several radar-based gestures of the deaf sign language using Short-time Fourier Transform (STFT) [15], however, the recognition of these gestures using deep learning has not been reported.

One of the Deep learning techniques that have been used to classify gesture signals is the Convolution Neural Network (CNN) [16-18]. Cuma et al [16] employed CNN to recognise static and dynamic gestures obtained from radar and have achieved an accuracy of 93.2% and 90.5% respectively in classifying the gestures. Detection of cough, stopping, moving arms, scratching head, and shaking head gestures have been investigated by [17] using radar and CNN. In their work, gestures from various distances; one-meter, three-meter, and mixed distances were studied. It was discovered that an accuracy of 88%, 80%, and 86.5% was achieved in detecting gestures from one meter, three meters, and mixed distances. To the best of our knowledge, CNN has not been used to classify gestures of deaf sign language.

This paper describes the recognition of six gestures signals of deaf sign language using radar and CNN. The recorded gestures signals were converted to images using STFT before passing through the classifier and CNN classification was performed using two approaches.

2. Method

The whole process of recognizing the radar-based deaf sign language is shown in Fig. 1. It includes hand gesture acquisition, gesture signal processing, feature extraction, segmentation, and classification of deaf sign language. The classification was performed on two types of input features; original and segmented spectrogram of gesture signals.

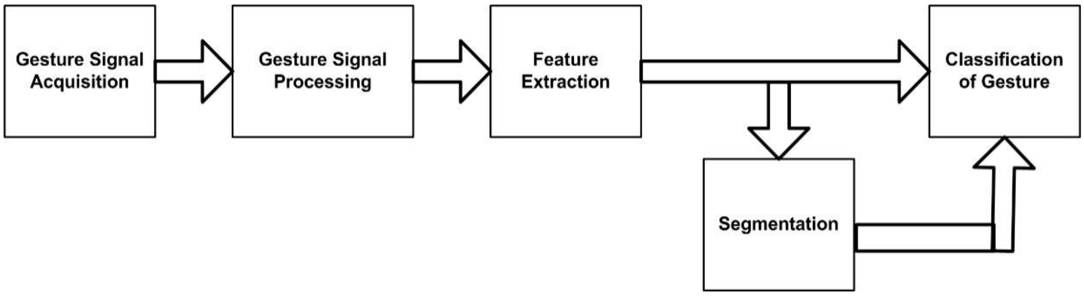

Fig. 1 - Block diagram of the whole process

2.1 Hand Gesture Acquisition

Six gestures of deaf sign language that are commonly used in deaf people communication were selected in this study. There are I, He, His, Want, Hello and Thank you as shown in Fig. 2. The gestures were recorded from twelve normal subjects aged 24 to 56 years using the ST100 radar starter kit. The sampling frequency used was 44kHz. In a sitting position in front of the radar system, each subject performed six gestures with twenty times repetition for each gesture. This yields a total of 1440 gesture signals. During the recording, their hands were at a distance of approximately 20cm from the radar as shown in Fig. 3. To differentiate between each repeated gesture signal, the subject paused for one second between every hand gesturing.

2.2 Gesture Signal Processing and Feature Extraction

Before feature extraction was performed, the gesture signals were first downsampled to 1 kHz. Since CNN works well with images, STFT was computed to convert the signal to images and produces spectrograms which serve as the input of the CNN. Equation (1) was used in the computation of the STFT. To get the high time and frequency resolution, the window size of a Hanning window was set to 64ms and the overlap length is 48ms. Thresholding was then performed to remove the background noise.

$$STFT\{y(t)\}(\tau,\omega) = \int_{-\infty}^{+\infty} y(t)\omega(t-\tau)e^{-jwt}dt \quad (1)$$

where $\omega(t)$ is the window function and $y(t)$ is the gesture signal.

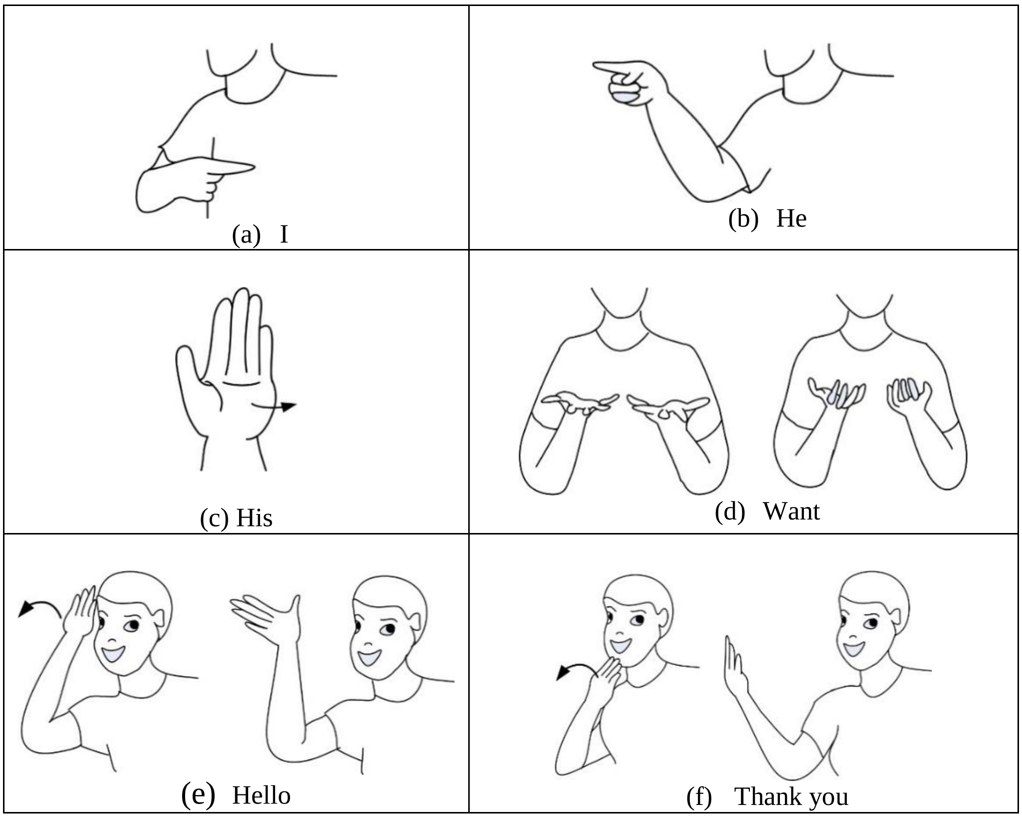

Fig. 2 - Hand gestures used in the study

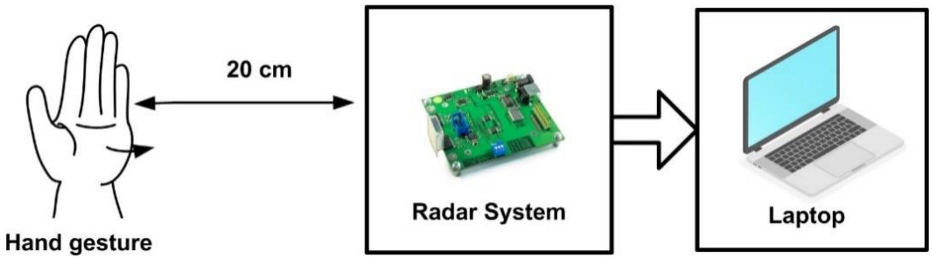

Fig. 3 - The arrangement of the system

2.3 Classification of Gesture

The CNN classifier utilized in this work comprises multiple layers. The most significant layers are the convolution layer or known as the kernel, pooling layer, fully connected layer, and classification layer. The convolution layer is responsible for extracting the high-level characteristics of the spectrogram of the gesture signals. The pooling layer is in charge of lowering the spatial size of the convolved feature and extracting the key features of the images. The fully connected layer connects all neurons from the layer to those of the previous layer. The last layer is the classification layer which produces the output of the classifier. In this layer, the soft-max loss function was used to predict the output and to produce the actual output. The CNN classifier was developed using GoogLeNet architecture and the Adaptive Moment Estimation (Adam) learning algorithm was used since it converges very fast. The data were divided into training, testing, and validation sets with the percentage of 80%, 10%, and 10% distribution respectively [19, 20]. In the training, the data was trained in a small batch. A heuristic optimisation approach was performed to identify the optimum values of the mini-batch size and maximum epoch. The mini-batch size and maximum epoch were varied in the range of 3 to 15, and 10 to 40 respectively. The Adam learning algorithm shown in (2) was employed to update the model parameters continuously during each training epoch to increase the training accuracy. The input size of the images used was 224 x 224.

$$\theta_{j+1} = \theta_j - \frac{\alpha m_j}{(n_j)^{\frac{1}{2}}+\epsilon} \tag{2}$$

where θ is the updated parameter, α is the learning rate, m is the first moment, n is the second moment and ϵ is a constant, usually 10^{-8}. The first and second moments were computed using (3) and (4) respectively.

$$m_j = \beta_1 m_{j-1} + (1-\beta_1)\delta E(\theta_j) \tag{3}$$

where β_1 is the first decay rate, E is the loss function which in this case is the soft-max and $\delta E\left(\theta_j\right)$ is the gradient which is obtained from the partial derivation of E with respect to θ.

$$n_j = \beta_2 n_{j-1} + (1-\beta_2)[\delta E(\theta_j)]^2 \tag{4}$$

where β_2 is the second decay rate.

The hand gestures were classified using two approaches. The first approach used the original spectrogram without applying segmentation as the input of the classifier. The second approach implemented segmentation by thresholding the spectrogram. The performance of the CNN model was then evaluated using a confusion matrix that displays four combinations of the predicted and actual values; true positive (T_p), true negative (T_n), false positive (F_p), and false negative (F_n). The classification accuracy (A_c), sensitivity (S_e), and specificity (S_p) were computed using (5), (6), and (7) respectively.

$$A_c = \frac{T_p+T_n}{T_p+T_n+F_p+F_n} \tag{5}$$

$$S_e = \frac{T_p}{T_p+F_n} \tag{6}$$

$$S_p = \frac{T_n}{T_n+F_p} \tag{7}$$

3. Results and Discussion

Fig. 4 shows the accuracy of CNN in recognising each hand gesture for various mini-batch size for training, validation, and testing. It is observed that the highest training accuracy (96.8%) is achieved when the mini-batch sizes of 9 and 15 are used. For both validation and testing accuracies, the mini-batch size of 12 produces the highest accuracy which is 86.39% and 88% respectively. Since the mini-batch size of 12 outperforms other values and produces high accuracy for training (94.23%), it is selected as the optimum mini-batch size and was used throughout the experiment. Besides, this mini-batch size offers fast computation time for the training compared to mini-batch sizes 3 to 9 (see Fig. 5).

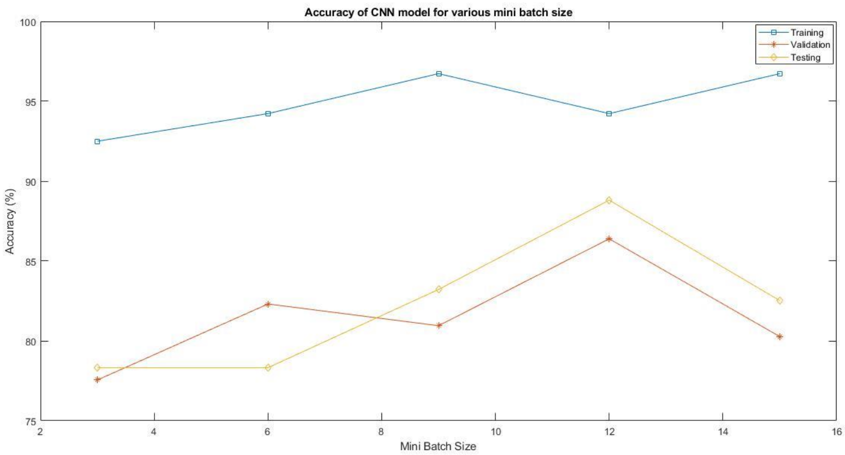

Fig. 4 - Training, validation and testing accuracies of CNN for various mini-batch size

The effect of varying the maximum epoch is shown in Fig. 6. Increasing the maximum epoch from 10 to 30 when the optimum mini-batch size is used causes the training accuracy to be increased. The highest training accuracy is 99.83% at a maximum epoch of 30 (see Fig. 6). The same trend is observed for the validation accuracy where the highest accuracy

(95.24%) is at a maximum epoch of 30. For the testing accuracy, it decreases at the maximum epoch of 20 and then increases until it reaches the highest testing accuracy which is 92.31% at the maximum epoch of 30. Since the maximum epoch of 30 produces the highest accuracy, it is selected as the optimum value. Table 1 shows the accuracy of recognising each hand gesture for testing data without segmentation. The model manages to distinguish the hand gestures with an average accuracy above 92.31%. The highest accuracy is achieved (96 %) when the CNN is classifying the 'Hello' gesture with sensitivity and specificity of 96%. The CNN produces the highest sensitivity (100%) in recognising Thank you and the highest specificity (100%) when distinguishing I.

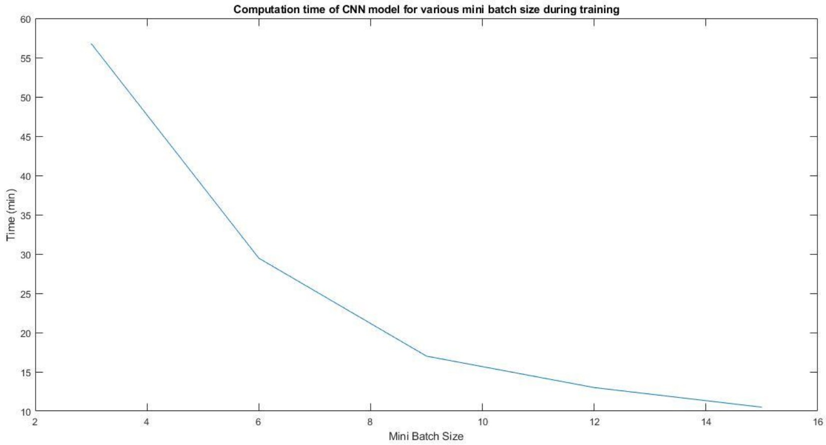

Fig. 5 - The computation time of the CNN model for various mini-batch sizes during training

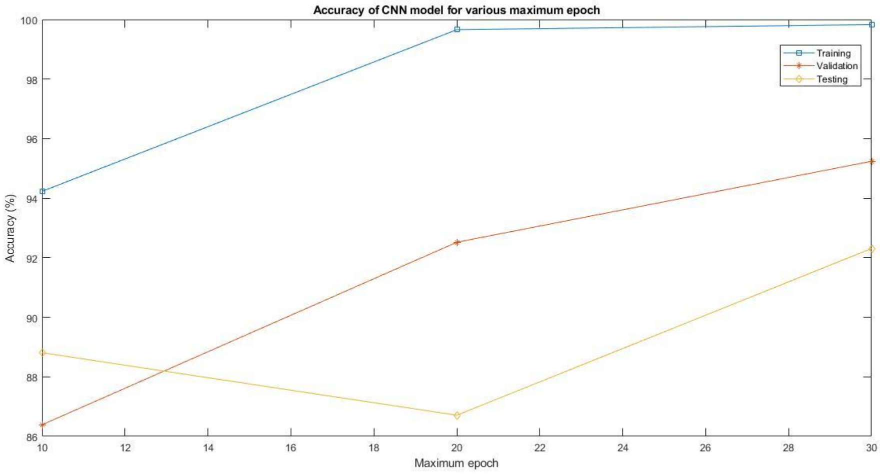

Fig. 6 - Training, validation and testing accuracies of CNN for various maximum epoch

When the segmentation is employed, the average accuracy of the CNN decreases to 90.9%. However, in recognising some of the gestures (He, His, Hello, and Thank you), the accuracy for testing data is higher than those without segmentation as shown in Table 1. The CNN manages to recognise He, His, Hello, and Thank you with accuracies of 94.40%, 95.85%, 96.15%, and 97.90% respectively. Using the segmented spectrogram, the highest sensitivity obtained is only 96.2% when He is classified and the highest specificity is 100% when Hello and Thank you are distinguished.

Table 1 - The testing accuracy of classifying each gesture with and without segmentation

Hand gesture	Accuracy (%)		Sensitivity		Specificity	
	Original	With Segmentation	Original	With Segmentation	Original	With Segmentation
I	95.85	82.40	91.70	84.00	100	80.80
He	92.15	94.40	95.80	96.20	88.50	92.60
His	81.70	95.85	73.90	91.70	89.50	100
Want	93.70	78.70	95.70	83.30	91.70	74.10
Hello	96	96.15	96	92.30	96	100
Thank you	94.45	97.90	100	95.80	88.90	100
Average Accuracy	92.31	90.90				

4. Conclusion

The recognition of gestures of deaf sign language using radar and CNN has been described in this paper. Conversion of gesture signals into images in the form of spectrograms was carried out to provide inputs to the CNN. The CNN was examined using spectrograms with and without thresholding to identify the optimal features. It was found that the CNN could recognise the gestures accurately and produce higher accuracy (92.31%) using non-segmented spectrograms of gesture signals than that of segmented spectrograms. Thus, the optimal method of recognising the radar-based gesture signals using CNN is without implementing segmentation on the features.

Acknowledgement

The authors would like to thank Microwave Research Institute, Universiti Teknologi MARA for providing equipment and support to carry out this work.

References

[1] Ming J C, Omar Z and Jawad M H 2019 A review of hand gesture and sign language recognition technique *International Journal of Machine Learning and Cybernetics* **10** 131.

[2] Phat N H and Tan P N Hand Gesture Recognition Algorithm Using SVM and HOG Model for Control of Robotic System 2021 *Journal of Robotics* 1.

[3] Mahmood M R, Abdulazeez A M and Orman Z A New Hand Gesture Recognition System Using Artificial Neural Network **2017** *International Scientific Conference-University of Zakho* 1.

[4] Kazllarof V, Karlos S and Panagopoulou A Automated hand gesture recognition for educational applications 2016 *Proceedings of the 20th Pan-Hellenic Conference on Informatics* 20 1.

[5] Juneja S, Juneja A, Dhiman G, Jain S, Dhankhar A, and Kautish S Computer Vision-Enabled Character Recognition of Hand Gestures for Patients with Hearing and Speaking Disability 2021 *Mobile Information Systems* 1.

[6] Bargellesi N, Carletti M, Cenedese A, Sust G A and Terzi M A Random Forest-based Approach for Hand Gesture Recognition with Wireless Wearable Motion Capture Sensors 2019 *IFAC PapersOnLine 52-111 128*

[7] Zhang G, Lan S, Zhang K, and Ye L Temporal-Range-Doppler feature interpretation and recognition of hand gestures using mmW FMCW radar sensors 2020 *14th European Conference on Antennas and Propagation* 7.

[8] Nguyen M Q, Flores-Nigaglioni A, and Li C Range-gating technology for millimeter-wave radar remote gesture control in IoT applications 2018 *IEEE MTT-S International Wireless Symposium, - Proceedings* 1.

[9] Lan S, He Z, Tang H, Yao K, and Yuan W A hand gesture recognition system based on 24GHz radars 2017 *International Symposium on Antennas and Propagation* 1.

[10] Shahzad Ahmed 1 , Karam Dad Kallu 2, Sarfaraz Ahmed 1 and Sung Ho Cho Hand Gestures Recognition Using Radar Sensors for Human-Computer-Interaction: A Review 2021 *Remote Sensing* 13 527.

[11] Hakobyan G. and Yang B High-Performance Automotive Radar: A Review of Signal Processing Algorithms and Modulation Schemes *IEEE Signal Processing Magazine 2019* **36** 5 32.

[12] Yilmaz B. and Karsligil M E Detection of Airplane and Airplane Parts from Security Camera Images with Deep Learning 2020 *28th Signal Processing and Communications Applications Conference Proceedings* 1.

[13] Sokol Z, Szturc J, Orellana-Alvear J, Popová J, Jurczyk A and Célleri R The Role of Weather Radar in Rainfall Estimation and Its Application in Meteorological and Hydrological Modelling A Review 2021 *Remote Sensing* 13 351.

[14] Rodriguez S. Wang P, Willems P, and Onof C A Review of Radar-Rain Gauge Data Merging Methods and Their Potential for Urban Hydrological Applications 2019 Water Resources Research 55 6356.

[15] Dol Malek M D H, Mansor W and Abdul Rashid N E Radar Based Gestures of Deaf Sign Language Signatures 2021 *IEEE Symposium on Wireless Technology & Applications* 93.
[16] Park J, Jang J, Lee G, Koh H, Kim C, and Kim T W A Time Domain Artificial Intelligence Radar System Using 33-GHz Direct Sampling for Hand Gesture Recognition 2020 *IEEE Journal of Solid-State Circuits* 55 879.
[17] Chuma E L and Iano Y a Movement Detection System Using Continuous-Wave Doppler Radar Sensor and Convolutional Neural Network to Detect Cough and Other Gestures 2021 *IEEE Sensors Journal* **21** 2921.
[18] Sun Y, Fei T, Li T X, Warnecke A, Warsitz E, and Pohl N Real-time radar-based gesture detection and recognition built-in an edge-computing platform 2020 *IEEE Sensors Journal* **20** 10706.
[19] Uçar M K, Nour M, Sindi H and Polat K, The Effect of Training and Testing Process on Machine Learning in Biomedical Datasets 2020 *Mathematical Problems in Engineering* 1
[20] Gholami E, Chau K W, Fadaee F, Torkaman J and Ghaffari A Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers 2015 *Journal of Hydrology* 1060.