Millimeter Wave Radar Sensing Technology for Filipino Sign Language Recognition

Jorelle Aaron Herrera, Almira Astrid Muro, Philip Luis Tuason III(✉), Paul Vincent Alpano, and Jhoanna Rhodette Pedrasa

University of the Philippines Diliman, Quezon City, Philippines
{jorelle.aaron.herrera,luis.tuason}@eee.upd.edu.ph, afmuro@up.edu.ph

Abstract. Filipino Sign Language (FSL) is the primary language used by the Deaf and Hard-of-Hearing (DHH) community in the Philippines. The lack of support for FSL from the government has led to a huge communication gap between the DHH and the hearing society. A substantial amount of research has been done to develop sign language recognition systems based on computer vision or wearable technology. However, most such systems are limited to controlled settings, while wearable systems also raise issues such as inconvenience to users. Millimeter wave (mmWave) technology has recently seen potential in gesture recognition applications as it allows the system to be non-contact and resistant to environmental factors while ensuring high resolution for recognizing small movements. An mmWave-based FSL recognition system that can translate isolated signs into their equivalent gloss was developed. Data from a TI IWR1443 radar sensor was fed into a preprocessing algorithm and a deep learning model composed of multi-view 2D CNNs and LSTM. 4 models were trained based on a dataset of 24 FSL signs gathered with 3 native Deaf signers in 3 different environments. A total of 3240 samples were collected, resulting in a model that attained an overall peak accuracy of 94.9% and an average real-time recognition latency of about 2.01 s. The model's performance is comparable to both existing FSL and mmWave systems, showing immense potential for future work on FSL recognition using mmWave.

Keywords: Sign Language Recognition · Millimeter Wave · Deep Learning

1 Introduction

Filipino Sign Language (FSL) is a complex visual language composed of forming hand shapes and movements mixed with non-manual signals such as facial

Supported by Philippine Federation of the Deaf.

A. Tsanas and A. Triantafyllidis (Eds.): PH 2022, LNICST 488, pp. 274–288, 2023.
https://doi.org/10.1007/978-3-031-34586-9_19

expressions and upper body movements. It is the primary mode of communication used by the Deaf and Hard-of-Hearing (DHH) community in the Philippines. FSL was initially based on the American Sign Language (ASL) but has since evolved to be an independent language with the continuous addition of new local signs with each generation of the Deaf [11].

As of 2009, 1.23% of the Philippine population is either deaf, mute, or hearing-impaired, 517,536 of whom have some degree of deafness [1]. For the past decades, the Deaf community has greatly lacked needed support from the government. Only recently was FSL recognized as the national sign language of the Philippines through R.A. 11106, also known as the FSL Act of 2018, which was a watershed moment for FSL and Deaf culture in the country [22].

Living in a hearing-centric society, the DHH community is exposed every day to several environmental vulnerabilities, which include communication barriers, additional disabilities, and a lack of mental health services. Furthermore, since the COVID-19 pandemic began, their need for social support has intensified as everyone started working and living remotely–mainly relying on the internet and technological devices, the content of which caters mostly to the hearing community. Interpreting languages is crucial for them to survive, and the lack of interpreters poses a demand for technology that can facilitate a relay service for the Deaf [21].

Efforts to build interpreting technologies have been going on for the past decades as a number of FSL recognition research has been conducted in the Philippines. The most used detection medium in past studies is computer vision through cameras [4,6] and infrared sensors like Microsoft Kinect [23] while other studies have tried using wearables such as gloves equipped with sensors for more precision [17]. More recent global developments have utilized radar sensing such as Wi-Fi signals and millimeter-wave (mmWave) signals for ASL recognition and other gesture recognition applications. This type of technology holds an advantage over cameras and wearables due to their limitations such as causing discomfort of wearing gloves, having limited range, requiring ideal lighting conditions, and raising potential privacy concerns [14].

Millimeter-wave signals have been found to show more potential than Wi-Fi signals due to their ability to detect finer movements and their robustness to environmental factors such as the movement of other people and objects. [32]. Furthermore, the recent global deployment of 5G gave rise to more research on technologies that it utilizes, such as mmWave frequencies. The future expansion of 5G networks will make mmWave infrastructures more accessible to the point where they will be collaborating with today's communication infrastructures, proving its huge potential in various applications [5].

2 Related Work

2.1 Filipino Sign Language

Filipino Sign Language (FSL) is a complete, natural visual language used by a majority of the Filipino Deaf community. FSL, much like all sign languages, has its own linguistic rules for pronunciation, word formation, and order. Much like how hearing persons have different ways of speaking, signers also express themselves differently. It comes with regional variations and dialects that differ down to the smallest but most significant parameters of a sign. Other sociological factors such as geographical location, age, and gender contribute to the variety and growth of sign language [18].

The fundamental structure of a sign mainly revolves around the model developed by Liddell and Johnson in 1989, describing the 5 parameters: hand shape, location, palm orientation, movement, and non-manual signals [13]. Hand shape pertains to the arrangement of the fingers and joints, while location refers to the position of the hands relative to the body [26]. Palm orientation refers to the direction the palm is facing, and movement can refer to the movement of the fingers or the path that the hand or arms take [26]. These 4 parameters make up the manual markers of sign language. The fifth parameter is composed of the non-manual signals. Some non-manual signals that have been recorded in FSL signs include even the smallest of movements and expressions on the face, such as brow and lip movements, eye gazes, and nose wrinkling [26].

2.2 Sign Language Recognition

Sign language recognition is generally classified into two approaches: glove-based and computer vision-based (CV-based) [30]. The main advantages of gloves are their higher accuracy and hand information extraction since the sensors are directly attached to the hand. However, among its limitations are its inability to provide other essential information, such as non-manual signals and movements in the rest of the body, and the inconvenience of wearing gloves, which may restrict the movement or expression of the signer [30]. The CV-based approach, on the other hand, has been widely researched due to the ubiquity of computer vision and the rapidly rising trends in machine learning and artificial intelligence [20]. However, it still comes with its limitations, such as its sensitivity to lighting conditions [6] and other environmental factors such as unwanted objects in the video [14].

2.3 Millimeter Wave

Millimeter wave (mmWave) refers to the spectrum between 30 and 300 GHz, which has wavelengths in the millimeter range (1 to 10 mm) [3,12,16]. This frequency range has been used to develop radar sensors that could measure range, velocity, and angle by transmitting electromagnetic waves and comparing them to the received reflections of those transmitted signals [12]. Because of its high frequency, mmWave radar sensor implementations have small and closely-spaced antennas, which allow favorable characteristics such as smaller component sizes, greater availability of bandwidth, lower mutual interference between radars, and higher spatial resolution over other radar sensing technologies [3,16].

One of the many applications of mmWave radar sensing technologies is sign language recognition. This technology has been particularly on the rise in this application because of its notable advantages over camera-based and wearable device-based implementations, which include non-intrusive, device-free, and environment-resilient sensing [16,25,32]. Recent trends focus on optimizing aspects that make it viable for real-world applications, which include real-time [15,25,29,32], person-independent [15], environment-resilient recognition [25,32].

2.4 Recognition Algorithms

Millimeter wave sensors can produce different types of datasets which researchers can utilize—point clouds [14,19,25,29], continuous range doppler image sequences or spectrograms [27,32,33], and raw vibrations [7]. Among these, the most common datasets used are in the form of point clouds, which are scatter plots in the euclidean three-dimensional space.

Preprocessing Algorithms. With the mmWave sensors removing many of the points reflected by static objects through the built-in static clutter removal algorithm CFAR, the data points are from dynamic objects detected by the sensor as well as scatterings and reflections from the environment. Some of these data points can be noise in the form of outliers, which need to be removed. Most studies that used point clouds as a dataset [25] applied DBSCAN [8] or an improvement of this algorithm [14,31] for their outlier removal.

Deep Learning Architectures. Both spatial and temporal properties of the data are essential to gesture recognition applications. Consequently, learning these two properties concurrently is fundamental to the architecture to be built for this specific project. Various deep learning architectures in gesture recognition studies with millimeter wave as the medium involve neural networks, from basic convolutional neural network (CNN) models [7,32,33] to improved novel networks [14,19,29], for learning feature representations and long short-term memory (LSTM) modules [25,27,31] for modeling signs over time. Of all the architectures, CNNs with or without LSTM modules are the most common deep learning algorithms used in mmWave gesture recognition regardless of data type [7,27,31–33].

3 System Design

The overall system is composed of the radar sensor connected to a local machine containing the recognition module and graphical user interface (GUI), as illustrated in Fig. 1. The sensor is initialized through the GUI, which establishes a serial connection. Then, the system captures raw data from a user performing a sign in front of the radar sensor. The captured data is then fed into the recognition module, starting with the preprocessing algorithm to clean the data. It is then fed through the deep learning model to classify the data and return the gloss of the sign. The gloss is then displayed on the local machine through the GUI.

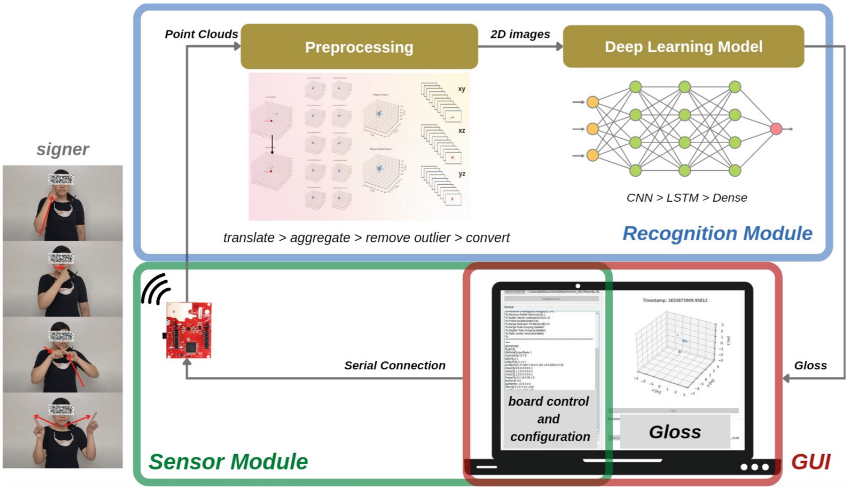


Fig. 1. Full system setup in real-time

3.1 Sensor Module

The sensor module used in this project is the IWR1443BOOST evaluation board which is a commercial off-the-shelf board from Texas Instruments that uses frequency-modulated continuous wave (FMCW) radars for high precision sensing at frequencies from 76 to 81 GHz.

Generation of Point Clouds. This board is a multiple-input and multiple-output (MIMO) device that makes use of range, elevation, and azimuth angle to generate point clouds in three-dimensional space [29]. The processes involved include Range-FFT (1D), Doppler-FFT (2D), Constant False Alarm Rate (CFAR), and Angle-FFT (3D) as shown in Fig. 2 [25, 29].

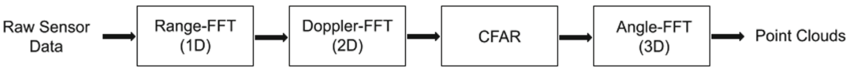


Fig. 2. Signal processing of IWR1443

Radar Configuration. The radar configuration file for the sensor module determines the characteristics of its transmitted signal and how it processes the signals it receives. The researchers configured the chirp properties to 20 fps, a velocity resolution (v_{res}) of 0.13 m/s, a range resolution (R_{res}) of 0.047 m, a maximum velocity (v_{max}) of 1.0 m/s, and a maximum range (R_{max}) of 2.41 m. In addition to this, the researchers also decided to turn off the Range Peak

Grouping and Doppler Peak Grouping, which reduce the number of data points by grouping those that are close together, to avoid the omission of necessary data points for feature extraction, and to turn on the Static Clutter Removal, which removes data points that are not in motion, to allow the system to be more resilient to noise caused by the multipath effects brought by static objects within the field-of-view of the sensor module.

3.2 Recognition Module

The recognition module is composed of two submodules, namely the preprocessing algorithm, which was inspired by the Pantomime study [25], and the deep learning model, which was derived from the ExASL study [31].

Preprocessing Module Algorithm. There are four stages in the preprocessing algorithm, as shown in Fig. 3. First, all points in all raw frames are translated to the origin. Once centered, the algorithm reduces the frames of each sample to the desired number by aggregating. Each of the aggregated frames is then subjected to noise reduction by removing outliers. Lastly, the 3D frames are converted into 2D multiview frames.

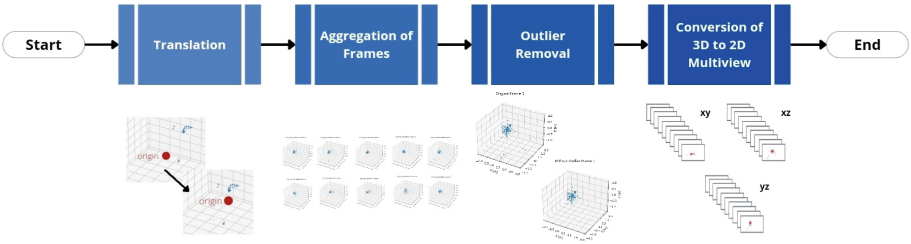


Fig. 3. Overview of the Preprocessing Algorithm

Translation. Raw point cloud data was translated into the main cluster based on the relative displacement of its centroid to the origin. At this stage, the position of the cloud is normalized, aiming to reduce the effect of the position of the signer.

Aggregation. To reduce the complexity of the module, the original number of frames of raw data from the sensor operating at 20 fps was reduced to 10 frames per sign. The method of aggregation used in this study is inspired by the time decay method of the Pantomime study [25]. All points were chronologically grouped into sets of k/f points, where k is the total number of points in the data and f is the desired number of frames per sign. In this study, the value of k varies from sign to sign while $f = 10$. Aggregating the frames resulted in a denser point cloud per frame, which enabled better outlier identification.

Noise Reduction. Point cloud data can contain noise due to scattering and reflections from the environment, especially in cluttered environments [25]. With the sensor having the function to remove points from static objects, noise from the sensor can be removed by simple outlier detection through the DBSCAN algorithm using $\epsilon = 0.5$ and minimum samples in a cluster ($min\_samp = 5$).

Conversion. The first stage of the layers of the deep learning model, CNN, required inputs of consistent and fixed dimensions. In addition, the model used 2D CNN, which takes in two-dimensional images. To meet these requirements, after aggregating the frames, the 3D point cloud data was converted into 2D multiview data. Subsequently, each sample has 3 view sets—xy-view, yz-view, and xz-view—and each view set has f frames.

Deep Learning Model. The structure of the model used in this study is derived from the study ExASL [31]. The model has three main components, which are the view-specific CNNs, LSTMs, and dense layers. The view-specific CNN portion contains 4 sets of (1) convolutional layers made of 5×5 convolutional kernels, (2) a max pooling layer with a 2×2 kernel, and (3) rectified linear units as activation arranged sequentially. Bidirectional LSTMs contain two layers of LSTM cells with 2048 hidden units. Lastly, the dense layer consists of three linear layers with 2048, 1024, and 512 hidden units, respectively. A dropout rate of $p = 0.65$ was also used for regularization [31].

3.3 Graphical User Interface

A simple graphical user interface (GUI) for the system was developed using Python 3 and PyQt5. The GUI was used in all stages of the project, namely interfacing the sensor with the local machine, facilitating the data-gathering process, and determining the feasibility of real-time applications of the system. It is able to display the console outputs of the board and displays the sensor's collected data through a 3D scatter plot. It is also used to load the deep learning model, allowing for the direct recognition of the signs.

3.4 Evaluation Parameters

Two parameters were considered for analyzing the performance of the system—sign-to-gloss accuracy (individual A_{S-G} and overall μ_A) and recognition latency (l_R). The individual sign-to-gloss accuracy, A_{S-G}, pertains to the accuracy of each of the 24 signs per model (Eq. 1), while the overall accuracy, μ_A, pertains to the accuracy of a model on all 24 signs (Eq. 2).

The recognition latency of the system, l_R, was measured by getting the mean time interval between the time when the data stream is passed from the sensor to the local machine (t_{SL}) and the time the user interface presents a final sign-to-gloss translation for the said sign (t_{UI}), with N being the total number of trials (Eq. 3).

$$A_{S-G} = \frac{\#\ of\ times\ G\ (gloss)\ is\ matched\ to\ this\ S\ (sign)}{total\ \#\ of\ attempts\ for\ S} \times 100\% \tag{1}$$

$$\mu_A = \frac{total\ \#\ of\ correct\ translations}{total\ \#\ of\ trials} \times 100\% \tag{2}$$

$$l_R = \frac{\sum_{n=0}^{N} t_{UI} - t_{SL}}{N} \tag{3}$$

4 Testing

4.1 Selection of Signs and Signers

A set of 24 signs was selected by Dr. Liza Martinez, the founder and former director of the Philippine Deaf Resource Center, and approved by the Philippine Federation of the Deaf (PFD). The selection was based on the signs' distinct features with respect to some basic characteristics of the phonological structure of a sign, namely the number of articulators, the location, the path, and the movement.

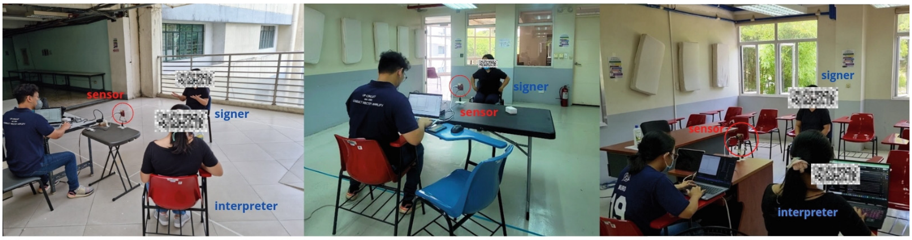
Three right-handed Deaf adults with similar signing styles volunteered, as coordinated with the PFD, as the project's participants to perform the signs in front of the system. In order to minimize potential variations in the manner of signing, the participants were selected to be native Deaf signers, meaning that they began signing in early childhood.

4.2 Physical Setups

This study defines noise as the multipath effects of static objects in the environment. To investigate the system's usability in a real-world setting and analyze the effect of the openness of an area on the signal propagation to and from the radar sensors, three different environments were implemented—(1) outdoor or open space (Fig. 4a), (2) indoor with minimal noise or an enclosed space with negligible background objects (Fig. 4b), and (3) indoor with noise or an empty classroom with static clutter (Fig. 4c). The order of pictures in Fig. 4 ranks the physical setups by noise exposure, from left having the least noise to right. In each setup, a chair was placed 1.5 m directly in front of the sensor.

4.3 Comparative Testing

Each of the three Deaf signers performed the 24 signs 15 times in each of the three environments, amounting to 1080 samples per environment and 3240 samples overall. This produced four datasets, one for each environment and one overall dataset containing all the samples from the three environments. These four datasets were used to train four individual models and were labeled as I for indoor with minimal noise, IWN for indoor with noise, O for outdoor, and C for combined.



(a) Outdoor (b) Indoor with Minimal Noise (b) Indoor with Noise

Fig. 4. Physical Setups

The four datasets were split into a ratio of 80:20 for training and testing sets, respectively. The models were trained with $batch\_size = 5$ and $epochs = 400$. An Nvidia RTX 3060 GPU with a 12 GB VRAM was used to train the models. After finalizing the models, all four of the test datasets were used to test each model's performance in terms of accuracy. The model with the highest accuracy was set to be the best model and was used in the real-time testing phase.

4.4 Real-Time Testing

The researchers conducted simple tests to determine the semi-real-time capability of the system. For logistical convenience, one of the researchers served as the signer and the setup was an outdoor environment located at the home of one of the researchers. Using the GUI and the best model achieved from comparative testing, the researchers recorded and tested 3 repetitions of each of the 24 signs. After each recording, the system automatically inferred the gloss corresponding to the performed sign with some delay. These delays were measured to compute the overall recognition latency of the system.

5 Results and Discussion

5.1 Performance of Models

Given that all test datasets were used to test each model, the performance of each model was analyzed by focusing on both the category testing results and cross testing results. This study defines *direct testing performance* as the performance of the models using the same test dataset used for training, i.e., the same environment, while *cross testing performance* is the performance of the models using datasets different from their respective categories.

Direct Testing Performance. Model O yielded the highest accuracy among the environment-separate models. This is consistent with the condition of the outdoor setup arranged in this study—an open space—which reduced the multipath effects of static objects such as walls in indoor setups. Comparing the three

Table 1. Overall Model Accuracies

Model	Accuracy (μ_A)
O	93.52%
I	88.89%
IWN	87.96%
C	**93.83%**

environment-separate models by accuracy as seen in Table 1, Model I yielded higher accuracy than Model IWN, but lower accuracy than Model O as expected based on the different levels of noise present in each setup. Therefore, these results show that the lower the noise level, the higher the model's accuracy.

Model C produced the highest accuracy (Table 1) among the four models since it was trained to classify glosses with 3 different levels of noise and 3 times the number of training samples processed than any other model. This suggests that the more data and variation used to train the model, the better it may translate signs accurately and the more resilient it can be to noise. Consequently, Model C was used in real-time testing where the researchers tested the system's real-time capability.

Ranking the environments with increasing noise levels, the outdoor setup comes first, followed by the indoor with minimal noise and the indoor with noise setup. The results in Table 2 are consistent with these setup conditions. Model I was trained using a dataset with little-to-no noise present, a noise level that can range from that of the outdoor dataset and the indoor with noise dataset. As a consequence, Model I yielded the highest accuracy and best cross testing performance among the other two models.

Table 2. Cross Testing Accuracies of Models

Model	**Test Dataset**			
	Outdoor	Indoor w/ Min. Noise	Indoor w/ Noise	Combined
O		56.94%	64.35%	71.60%
I	71.30%		72.22%	77.47%
IWN	62.96%	54.17%		68.36%
C	**94.91%**	**93.06%**	**93.52%**	

Cross Testing Performance. Model O yielded the best direct testing performance among the three models that are environment-separate, but the data used to train this model contains less noise than those of other setups. Hence, it performed the worst when tested with data with noise. This phenomenon is also evident in the cross testing results of Model IWN, with the most noise. Lastly, since Model C has the most exposure to different levels of noise, it yielded the best cross testing performance among all models (Table 2).

5.2 Analysis of Signs

While the overall accuracy of the models describes the overall performance of the model, individual sign-to-gloss accuracies describe how the model behaves with respect to each sign. Analyzing this metric provides insight into how well the model can read into the phonological features of the sign that were highlighted in this project. The average individual sign-to-gloss accuracy across all the models was measured to be 90.82%, showing that each model is sufficiently able to recognize and distinguish all 24 signs from one another. For further analysis, the signs were grouped according to their individual sign-to-gloss accuracy and the common features within such groups were identified in Table 3.

Since the mmWave radar sensor collects sparse point clouds as data, the system primarily relies on location and movement for recognizing signs. Signs that have a high recognition rate in the dataset are those with more straightforward or distinct paths and sequences of movements. On the other hand, the signs with similar small movements, such as twisting of the wrist and moving of fingers, have poorer performance. Nevertheless, the limited vocabulary of the system also limits the conclusions that can be drawn based on the structure of the signs. FSL, like all sign languages, is a very complex language whose lexicon continues to evolve today. These mispredictions made by the system can be attributed to other factors such as the environment, the signers, or the size of the datasets.

Table 3. Individual Sign-to-Gloss Accuracy of Model C

Sign to Gloss Acc (A_{S-G})	Glosses	Common features between at least 2 signs within the group
100%	FEEL_LAZY, MRT_LRT, PRETEND, COVID19, CONCLUSION, PINEAPPLE, EYE_EYE_DIFFERENT	Mostly one handed Single and straight paths; Multiple movement-hold segments
96.30%	MAID, WHERE, AGREE, UTANG, RENT, I_VISIT_YOU	Same path; Similar position of dominant and non-dominant hands
92.59%	18, WALA_PERA, LOLO_LOLA, YEAR	Twisting of wrist; Similar handshape
88.89%	YES, SAME, EXPOSE, OBSERVE,	Similar position and movement of both hands
81% to 86%	CIVIL_MARRIAGE, ROOF_TWIST, COUNT	Similar finger internal movement

5.3 Feasibility of Real-Time Implementation

For the real-time testing, a total of 72 samples were recorded, 3 repetitions of each of the 24 signs, performed by one of the researchers as the signer. The average recognition latency was measured to be 2.0086 s. The best performing model, Model C, and a laptop with a 3.1 GHz Dual-Core Intel Core i5 CPU and without a GPU were used for this test. In addition, out of the 72 signs recorded, 49 were inferred correctly by the system, which resulted in an accuracy of 68.06%. The significantly lower accuracy measured in this test can be attributed to the signing experience of the signer. The dataset used to train the model was made with native Deaf signers, while a beginner signer did this test.

5.4 Comparison with Existing Studies

Table 4 shows the best accuracy of this project compared with that of relevant works in FSL recognition and mmWave systems. This project is shown to outperform much of the previous work in FSL and it is comparable with that of other mmWave systems.

Table 4. Comparison with Results of Related Work

FSL Projects		mmWave Projects	
Project and Focus	Best Accuracy	Project and Data Type	Best Accuracy
mmWave FSL (this project)	**94.91%**	[2], Spectrogram	95.04%
[6] Isolated signs	89.00%	[9], 3D Point Cloud	95.00%
[20] Static FSL numbers	83.10%	[10], Spectrogram	72.50%
[23] Basic FSL signs using arms	95.00%	[25], 3D Point Cloud	96.12%
[24] Alphabet, numbers, 30 words	79.44%	[31], 3D Point Cloud	92.50%
[28] Facial expressions	76.00%	[32], Spectrogram	86.7%

Comparison with FSL Recognition Systems. According to Cabalfin et al. [6], limiting the signs used to a small number of very distinct features results in higher accuracy. This is reflected in this project, where the signs that are distinct and straightforward tend to be recognized more accurately. In contrast, those with similarities in a number of FSL phonological parameters are less reliably recognized. In particular, this project struggles with signs with similar positions and movement, which is avoided by Oliva et al. [23] by only using signs that can be performed with the use of the arms only, and with open-palm hand shapes. This may explain why its accuracy is as high as 95% even without using deep learning models, which generally have higher accuracies compared to those that use supervised machine learning techniques like SVM [32].

Comparison with Other MmWave Recognition Systems. The results from *Pantomime* [25] show that the highest accuracy model is trained in the open area and office setups and is tested in the open area. This is reflected by this project, as the best accuracy was achieved when the model was trained on the combined dataset and tested on the outdoor dataset, which is the most open setup.

The projects, *mmASL* [32] and *ASL Recognition Using RF Sensing* [10], tested the impact of different signers on the performance of models. [10] proved that the performances of models trained and tested on native signers and non-native signers are significantly different, while [32] observed that particular samples where the participants performed the sign slower than average produced errors. These observations were also reflected in the results of this study, particularly in the real-time testing phase.

6 Conclusion and Future Work

A functional recognition system for FSL was implemented using an mmWave FMCW radar sensor and deep learning. The sensor module collected signer data in the form of point clouds, and a preprocessing algorithm converted the point clouds into multi-view 2D images. The CNN and LSTM-based deep learning models, derived from ExASL [31], were trained and tested using the collected data. A graphical user interface was used to collect data and display the output point clouds, predicted gloss, and latency of prediction.

A dataset containing 3240 total samples of 24 signs performed by three (3) native FSL signers was used to train four (4) models (I, IWN, O, C) which corresponded to the different test environments. Model C, which had the most exposure to different levels of noise and the largest number of samples in its train dataset, achieved the best overall accuracy, best direct testing performance ($\mu_A = 93.83\%$), and best cross testing performance, which ranged from 93.06% to 94.91%. Consequently, Model C was used to test the real-time capability of the system, which yielded a recognition latency of ≈ 2.01 s. The findings are consistent with that of the existing FSL and mmWave recognition systems, and the model performance is competitive when compared to the same projects, especially in FSL recognition.

Improvements can be made in the various aspects of the study for future iterations. Introducing more variations, namely the number of signs, signers, and different environments to produce a more representative dataset. Furthermore, aside from point clouds, the radar sensor is also capable of producing micro-doppler spectrograms. Other studies [2,10,32,34] have successfully utilized this data type thus, future work can explore the combination of both data types. The preprocessing algorithm and the deep learning model also have parameters that can be varied and explored, in addition to trying entirely different model architectures such as 3D CNNs or Transformers. Finally, expanding the scope to translating sentences and conversations is most desirable in the near future.

References

1. Senate Bill No. 2117, An act requiring the use of Filipino Sign Language insets for local news programs, amending for the purpose Section 22 of Republic Act No. 7277, as amended, otherwise known as the Magna Carta for Persons with Disabilities (PWDs) (2014). https://legacy.senate.gov.ph/lisdata/1868815815!.pdf

2. Adeoluwa, O., Kearney, S., Kurtoglu, E., Connors, C., Gurbuz, S.: Near real-time ASL recognition using a millimeter wave radar, p. 43 (2021). https://doi.org/10.1117/12.2588616
3. Al-Hourani, A., et al.: Chapter 7 - millimeter-wave integrated radar systems and techniques. In: Chellappa, R., Theodoridis, S. (eds.) Academic Press Library in Signal Processing, vol. 7, pp. 317–363. Academic Press (2018). https://doi.org/10.1016/B978-0-12-811887-0.00007-9, https://www.sciencedirect.com/science/article/pii/B9780128118870000079
4. Balbin, J.R., et al.: Sign language word translator using neural networks for the aurally impaired as a tool for communication. In: 2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), pp. 425–429. IEEE (2016)
5. van Berlo, B., Elkelany, A., Ozcelebi, T., Meratnia, N.: Millimeter wave sensing: a review of application pipelines and building blocks. IEEE Sens. J. **21**, 10332–10368 (2021)
6. Cabalfin, E.P., Martinez, L.B., Guevara, R.C.L., Naval, P.C.: Filipino sign language recognition using manifold projection learning. In: TENCON 2012 IEEE Region 10 Conference, pp. 1–5. IEEE (2012)
7. Dong, Y., Yao, Y.D.: Secure mmWave-radar-based speaker verification for IoT smart home. IEEE Internet Things J. **8**(5), 3500–3511 (2021). https://doi.org/10.1109/JIOT.2020.3023101
8. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of 2nd International Conference on Knowledge Discovery, pp. 226–231 (1996)
9. Gurbuz, S., et al.: American sign language recognition using RF sensing. IEEE Sens. J. **21**, 3763–3775 (2020). https://doi.org/10.1109/JSEN.2020.3022376
10. Gurbuz, S.Z., et al.: American sign language recognition using RF sensing. IEEE Sens. J. **21**(3), 3763–3775 (2021). https://doi.org/10.1109/JSEN.2020.3022376
11. Hurlbut, H.M.: Philippine signed languages survey: a rapid appraisal (2008)
12. Iovescu, C., Rao, S.: The fundamentals of millimeter wave radar sensors. https://www.ti.com/lit/wp/spyy005a/spyy005a.pdf
13. Liddell, S.K., Johnson, R.E.: American sign language: the phonological base. Sign Lang. Stud. **64**(1), 195–277 (1989)
14. Liu, H., et al.: Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 4, pp. 1–28 (2020)
15. Liu, H., et al.: M-gesture: person-independent real-time in-air gesture recognition using commodity millimeter wave radar. IEEE Internet Things J. **9**, 3397–3415 (2021). https://doi.org/10.1109/JIOT.2021.3098338
16. Liu, H.: Chapter 5 - autonomous rail rapid transit (art) systems. In: Liu, H. (ed.) Robot Systems for Rail Transit Applications, pp. 189–234. Elsevier (2020). https://doi.org/10.1016/B978-0-12-822968-2.00005-X
17. Magistrado, J.: Gloves na kayang mag-convert ng ph sign language sa boses, binuo ng ilang estudyante (2021). https://news.abs-cbn.com/news/07/02/21/sign-language-gloves-students-camsur
18. Martinez, L.B.: Observations on regional variants and handshape patterns of six signs in Filipino sign language (2009)
19. Meng, Z., et al.: Gait recognition for co-existing multiple people using millimeter wave sensing, vol. 34, pp. 849–856 (2020). https://ojs.aaai.org/index.php/AAAI/article/view/5430

20. Montefalcon, M.D., Padilla, J.R., Llabanes Rodriguez, R.: Filipino sign language recognition using deep learning. In: 2021 5th International Conference on E-Society, E-Education and E-Technology, pp. 219–225 (2021)
21. Movido, A.: Feeling left out, deaf community seeks government help to adjust to new normal (2020). https://news.abs-cbn.com/life/05/20/20/feeling-left-out-deaf-community-seeks-government-help-to-adjust-to-new-normal
22. Notarte-Balanquit, L.A.: Insights from the first Filipino Sign Language (FSL) summit & the prospects for Filipino sign linguistics (2021)
23. Oliva, K.E., Ortaliz, L.L., Tobias, M.A., Vea, L.: Filipino Sign Language recognition for beginners using kinect. In: 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), pp. 1–6. IEEE (2018)
24. Ong, C., Lim, I., Lu, J., Ng, C., Ong, T.: Sign-language recognition through gesture & movement analysis (SIGMA). In: Billingsley, J., Brett, P. (eds.) Mechatronics and Machine Vision in Practice 3, pp. 235–245. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76947-9_17
25. Palipana, S., Salami, D., Leiva, L.A., Sigg, S.: Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **5**(1), 127 (2021). https://doi.org/10.1145/3448110
26. Philippine Deaf Resource Center, I., Philippine Federation of the Deaf, I.: An Introduction to Filipino Sign Language, vol. 1: Understanding Structure. Philippine Deaf Resource Center, Inc. (2004)
27. Ren, Y., Lu, J., Beletchi, A., Huang, Y., Karmanov, I., Fontijne, D., Patel, C., Xu, H.: Hand gesture recognition using 802.11ad mmWave sensor in the mobile device. In: 2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), pp. 1–6 (2021). https://doi.org/10.1109/WCNCW49093.2021.9419978
28. Rivera, J.P., Ong, C.: Recognizing non-manual signals in Filipino Sign Language. In: Proceedings of Eleventh International Conference on Language Resources and Evaluation (LREC 2018), pp. 1–8 (2018)
29. Salami, D., Hasibi, R., Palipana, S., Popovski, P., Michoel, T., Sigg, S.: Tesla-rapture: a lightweight gesture recognition system from mmWave radar point clouds (2021)
30. Sandjaja, I.N., Marcos, N.: Sign language number recognition. In: 2009 Fifth International Joint Conference on INC, IMS and IDC, pp. 1503–1508. IEEE (2009)
31. Santhalingam, P.S., et al.: Expressive ASL recognition using millimeter-wave wireless signals. In: 2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), pp. 1–9 (2020). https://doi.org/10.1109/SECON48991.2020.9158441
32. Santhalingam, P.S., Hosain, A.A., Zhang, D., Pathak, P., Rangwala, H., Kushalnagar, R.: mmASL: environment-independent ASL gesture recognition using 60 GHz millimeter-wave signals. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 1, pp. 1–30 (2020)
33. Wang, J., Ran, Z., Gao, Q., Ma, X., Pan, M., Xue, K.: Multi-person device-free gesture recognition using mmwave signals. China Commun. **18**(2), 186–199 (2021). https://doi.org/10.23919/JCC.2021.02.012
34. Wang, Z., Yu, Z., Lou, X., Guo, B., Chen, L.: Gesture-radar: a dual doppler radar based system for robust recognition and quantitative profiling of human gestures. IEEE Trans. Hum. Mach. Syst. **51**(1), 32–43 (2020)