Sign Language Vocabulary Recognition Only with Tactile Sensing Glove

Motoki Kagami

School of Computer Science and Engineering,
The University of Aizu,
Aizu-Wakamatsu, Japan
Email: kagami.motoki.ki3@naist.ac.jp

Zeping Yu

Graduate School of Computer Science
and Engineering,The University of Aizu,
Aizu-Wakamatsu, Japan
Email: d8251103@u-aizu.ac.jp

Sim Teck Ceng

Department of Management and Design,
University of Aizu Junior College Division,
Aizu-Wakamatsu, Japan
Email: tcsim@jc.u-aizu.ac.jp

Lei Jing

Graduate School of Computer Science
and Engineering, The University of Aizu,
Aizu-Wakamatsu, Japan
Email: leijing@u-aizu.ac.jp

***Abstract*— A significant challenge in communication arises between deaf individuals and hearing individuals. Sign language recognition is expected to be an important research field for eliminating barriers. In this field, data collection is primarily conducted using two methods: sensor-based and vision-based approaches. In this study, we adopted a sensor-based approach by using a tactile sensing glove due to its portability, cost-effectiveness, and ability to capture fine finger movements. Subsequently, we applied LSTM and k-NN respectively to evaluate the recognition accuracy using this method. As a result, LSTM achieved a test accuracy of approximately 76%, while k-NN demonstrated an average accuracy of 87%. These findings highlight the potential of sensor-based SLR technology in reducing communication barriers and advancing inclusively in communication tools.**

I. Introduction

According to the World Health Organization (WHO), more than 5% of the world's population, equivalent to more than 430 million people, requires rehabilitation for hearing loss. This population includes deaf and hard-of-hearing individuals who rely on sign language as a vital means of communication. Sign language, which uses hand gestures, facial expressions, and body movements, is indispensable for communication within these communities. However, sign language remains largely unfamiliar to the hearing community, creating unresolved communication barriers between them and the deaf community. Sign Language Recognition (SLR) technology holds promise to overcome these barriers by facilitating communication and promoting social inclusion.

In Japan, the Ministry of Health, Labor and Welfare reports that there are approximately 340,000 individuals with hearing impairments, predominantly using Japanese Sign Language (JSL) and Signed Exact Japanese (SEJ). Although much of the existing research has focused on American Sign Language (ASL) [1], this study aims to advance SLR technology by emphasizing JSL, thereby promoting diversity and inclusivity. By focusing on JSL, we hope to address the unique challenges and nuances specific to Japanese sign language users, which are often overlooked in global research.

SLR research encompasses various data collection methods, primarily categorized into sensor-based and vision-based approaches. Sensor-based methods use devices such as accelerometers and bend sensors [2], while vision-based methods rely mainly on cameras for data capture [3]. Vision-based methods, although widely used, are vulnerable to environmental factors such as noise and lighting, which can significantly affect accuracy and reliability [4]. In contrast, sensor-based approaches remain largely unaffected by these influences, making them a robust alternative for reliable sign language recognition.

Given this background, our research focuses on the recognition of sign language vocabulary using only a tactile sensing glove, which incorporates force-sensitive films and conductive threads. Our tactile sensing glove is inexpensive and portable, and it can capture the nuanced hand movements and gestures of sign language without relying on visual data, providing a more resilient and adaptable solution. This study will investigate the recognition rate and accuracy of this approach, aiming to demonstrate its potential as a practical tool for real-world applications.

Through this research, we intend to contribute to the development of more inclusive and effective SLR technologies, ultimately bridging the communication gap between the hearing and deaf communities. Our findings could pave the way for further innovations in tactile sensing and its application in various assistive technologies, enhancing the quality of life for individuals with hearing impairments.

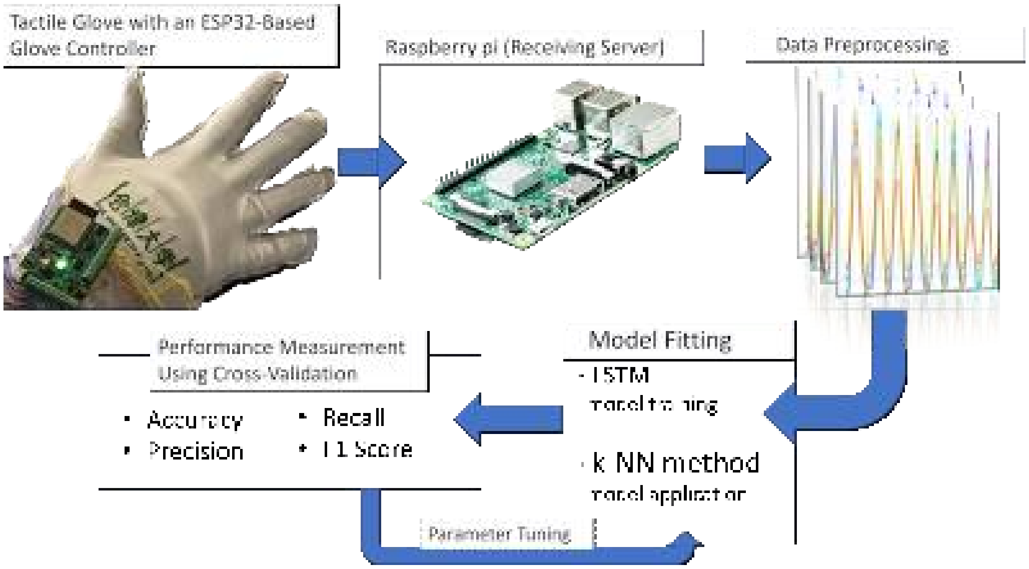


Fig. 1: The System Overview

II. Related Work

In the field of Sign Language Recognition (SLR), both vision-based and sensor-based methodologies have demonstrated their capability to automate the interpretation of sign language. Vision-based approaches primarily utilize camera footage to capture hand gestures and facial expressions. For example, the study by S.A.M.A.S. Senanayake et al. [3] combined camera footage with the MediaPipe framework for feature extraction and employed an LSTM network to process time series data, achieving training and validation accuracies of 94% and 90%, respectively. Additionally, Qizhi Gao et al. [5] used Kinect depth cameras to obtain RGB-D images of sign language. They utilized the SD-Segment image segmentation algorithm to align and segment images, and combined a dual-path feature blending attention network (DFANet) with a depth-pixel aware module (DPAM), achieving a maximum accuracy of 98.16

On the other hand, sensor-based strategies have shown promise in recognizing sign language with various sensor devices. Deemah Alosail et al. [2] explored the recognition of American Sign Language (ASL) and Arabian Sign Language (ArSL) using gloves equipped with bend sensors and accelerometers, achieving classification accuracies of 99.7% for ASL and 99.8% for ArSL using a Random Forest classifier. This study demonstrated that accelerometers contributed more significantly to model performance improvement than bend sensors. Additionally, a study on a magnetic sensor-based sign language recognition system [6] developed a system using six magnetic sensor nodes to measure the orientation of fingers and palms. By processing the measured orientation data with a deep learning classification algorithm, the system achieved nearly 100% classification accuracy for 26 sign language alphabets.

Furthermore, Subramanian Sundaram et al. [7] developed an inexpensive and portable tactile glove using pressure-sensitive films and conductive threads to measure the pressure distribution on fingers and palms at high resolution.

Unlike vision-based approaches, our study adopts a sensor-based approach that does not suffer from performance degradation due to environmental factors. Based on this prior work, we have developed a tactile glove with a specific focus on sensor placement. We strategically positioned the sensors at joint locations to capture more nuanced finger movements, enhancing the sensitivity and accuracy of our tactile glove.

A. *Design of Tactile Sensing Gloves*

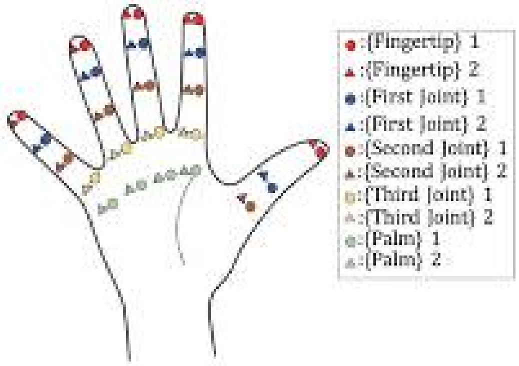
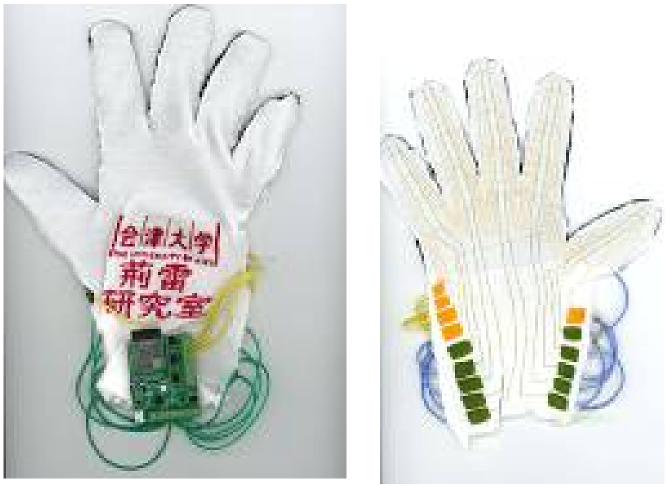


Fig. 2: Placement of the sensors

The tactile glove used in this study was fabricated according to the methods described in [7] and the glove is illustrated in Figure 3a, 3b.

The glove incorporates a total of 46 pressure-sensitive sensors made from ”Velostat,” a force-sensitive film and conductive threads, as illustrated in Figure 2. To precisely detect the movements of the fingers, sensors were strategically placed



(a) The Back Side of the Glove (b) The Front Side of the Glove

Fig. 3: The tactile sensing glove

on the fingertips, the joints of the fingers and the palm of the hand.

The tactile glove was mainly constructed using Brother's automatic embroidery machine, PRT5201, and the dedicated embroidery design software, Sisyu Pro 11, for designing the circuitry with conductive threads. The wiring design of the tactile glove used in the actual experiment with the software is shown in Figure 4.

B. Design of the Glove Controller

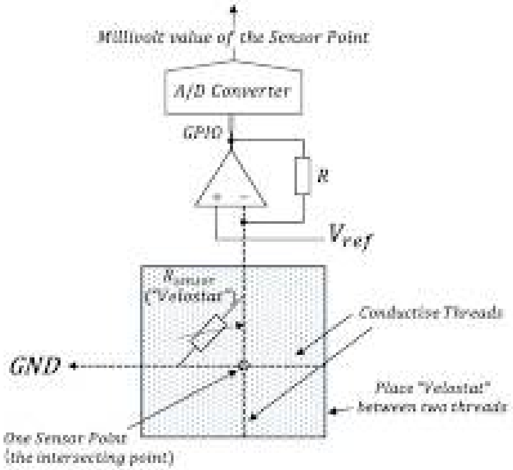

Millivolt value of the Sensor Point
A/D Converter
GPIO
R
V_{ref}
R_{sensor} ("Velostat")
Conductive Threads
GND
Place "Velostat" between two threads
One Sensor Point (the intersecting point)

Fig. 5: The Overall Circuit Design

$$V_{out} = \left(1 + \frac{R}{R_{sensor}}\right) * V_{ref} \tag{1}$$

The configuration of the pressure-sensitive sensor and its operating principle are shown in Figure 5,**??**. For the construction of the non-inverting amplification circuit, a fixed resistor of $5\,\mathrm{k\Omega}$ and a tactile sensor acting as the variable resistor R_{sensor} were utilized. A reference voltage V_{ref} of $0.33\,\mathrm{V}$ was applied, and the output voltage of this circuit is crucial for interpreting the sensor data, calculated using Equation (1).

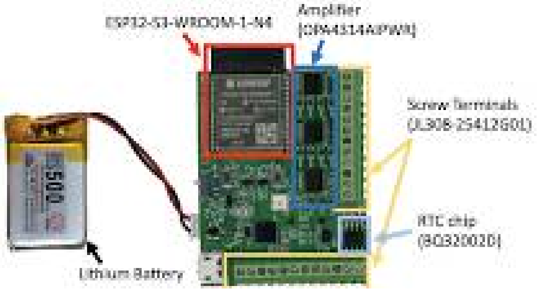


Fig. 6: The Glove Controller

The glove controller, as illustrated in Figure 6, was developed to process sensor data and transmit the data to the receiving server. It is equipped with an ESP32-S3-WROOM-1, featuring 20 ADC channels and 45 GPIO pins. Data transmission is facilitated through the Wi-Fi module integrated into the ESP32. Additionally, the controller includes 3 quad-channel operational amplifiers (Texas Instruments OPA4314AIPWR) and a real-time clock (RTC) chip (BQ32002D) from the same manufacturer.

III. Sign Language Recognition

A. Dataset

.

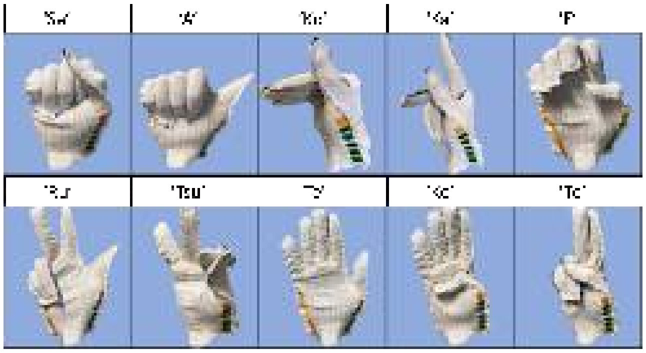


Fig. 7: The Sign Language Movements Used In This Study

In this study, we focus on the finger spelling used in SEL. The finger spelling represents Japanese syllables through specific hand shapes and positions, corresponding to each sound in the Hiragana script. This system serves as a supplementary means to the spoken Japanese language, playing a crucial role in communication for people with hearing impairments and those who use the sign language. Specifically, this research examines ten distinct finger characters, which correspond to the sounds 'Sa', 'A', 'Ko', 'Ka', 'E', 'Ru', 'Tsu', 'Te', 'Ke', and 'To'. Each sign language movement is illustrated in Figure 7.

In this investigation, sign language data was collected from five participants. Each participant performed each sign 25 times, resulting in a data set that comprises a total of 1,250 instances. This data set is constructed as follows: 5 participants × 25 repetitions per sign × 10 signs = 1,250 data points.

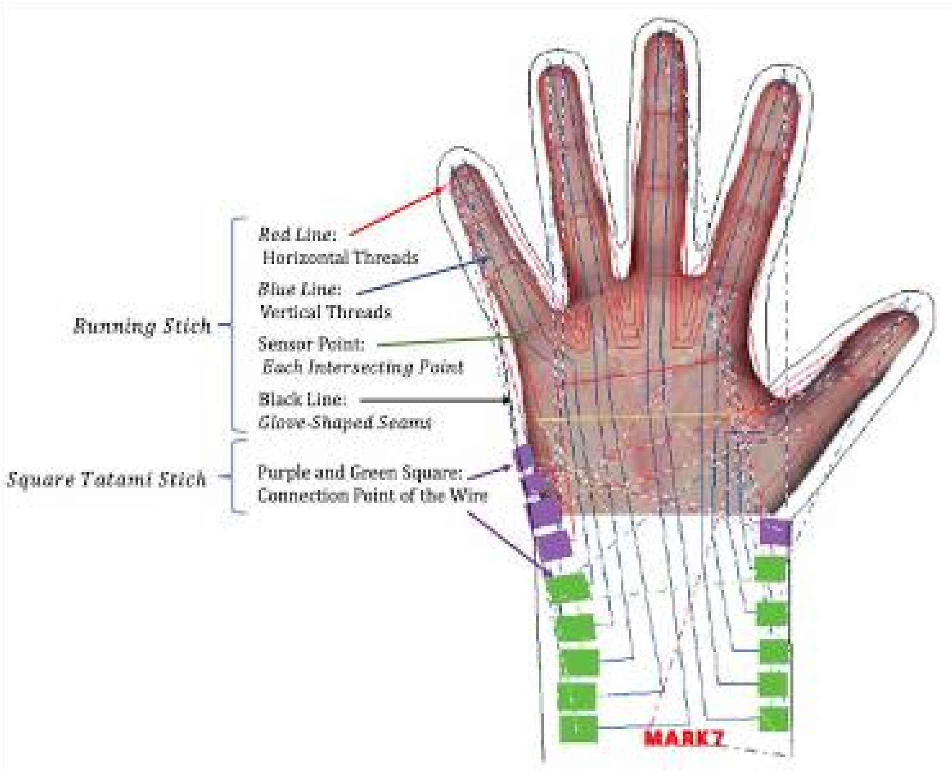


Fig. 4: The Wiring Design for Conductive Threads

B. Group 5-Fold Cross-Validation

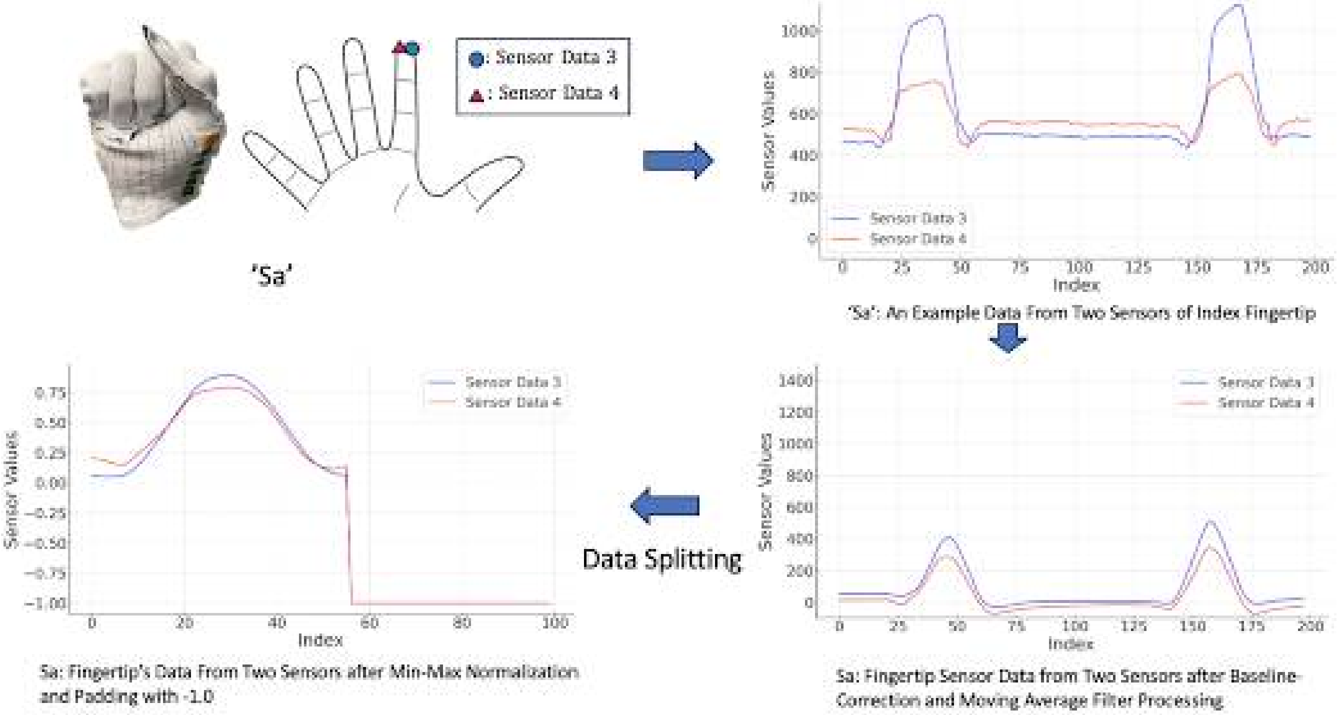
For the evaluation of our classification models, we employed a group 5-fold cross-validation method. In the case of the LSTM model, each cross-validation cycle involved using data from one participant as the test set, data from another participant as the validation set, and data from the remaining three participants for training, conducting a total of 20 permutations to ensure that each participant's data was assigned to the test, validation, and training sets at least once. Conversely, for the k-NN model evaluation, we utilized group 5-fold cross-validation without a validation set, selecting one participant's data as the test dataset and the data from the others for training in each cycle. This process was executed in 5 permutations, guaranteeing that each participant's data served as the test dataset once. Therefore, we comprehensively evaluated the data of each participant using both the LSTM and the k-NN models.

C. Sign Language Recognition Using LSTM and k-NN

In this investigation, our objective is to independently examine the use of LSTM and k-NN as classification methods for SLR, with the objective of evaluating their performance. The LSTM model is particularly skilled at handling dependencies over time, which benefits the understanding of continuous patterns in sign language movements. However, given that our dataset focuses primarily on static signs with fewer temporal variations and is limited in size, we also consider the k-NN method to be a suitable option for SLR due to its simplicity and effectiveness in classification. Faced with the risk of overfitting due to the limited data, we apply dropout and L2 regularization techniques in the training of the LSTM model using Keras to aim for better generalization.

Additionally, in k-NN, we facilitate the learning of time series data using the 'tslearn.neighbors.KNeighborsTimeSeries' class. This class employs Dynamic Time Warping (DTW) to adjust for temporal shifts in time-series data, enabling effective comparison of datasets that may differ in timing but are similar in shape. In this process, DTW also functions as a critical metric for measuring similarities among time series data within the 'KNeighborsTimeSeries' class. This functionality is essential to accommodate individual differences in the execution speed of sign language. Using DTW as both a key indicator and a measure of similarity, our k-NN model is further enhanced to intuitively and efficiently classify time series data, thus deepening our analysis of Sign Language Recognition (SLR).

Ultimately, we evaluate these methods using metrics such as accuracy, precision, recall, and the F1 score through group 5-fold cross validation, allowing us to quantitatively compare the effectiveness of each model in the sign language recognition task. This approach enables us to identify the most suitable model for SLR through a quantitative assessment of performance metrics.



'Sa': An Example Data From Two Sensors of Index Fingertip

Sa: Fingertip Sensor Data from Two Sensors after Baseline-Correction and Moving Average Filter Processing

Sa: Fingertip's Data From Two Sensors after Min-Max Normalization and Padding with -1.0

Fig. 8: The process of the preprocessing

IV. Experiments

A. Data Aquisition

Data were transmitted from the glove controller to the Raspberry Pi via Wi-Fi through TCP communication at a frequency of 100Hz in binary format. Each data point includes the device number, sensor number, epoch time with millisecond timestamp, data validation status, and values for each sensor data.

In the data collection phase, we initially invited five participants as previously mentioned. During the experiment, we collected data categorized by types of sign language movements. In addition, participants were instructed to ensure that each sign language movement lasted approximately 2 seconds from start to finish. For subsequent data segmentation, the experiment was recorded on video to individually segment each sign language movement.

B. Data Preprocessing

1) The issue with the Data: Upon collecting data using the fabricated glove, it was observed, as illustrated in Figure 9, that the steady-state values of the sensors varied significantly from one sensor to another. This discrepancy is attributed to the stress exerted on the sensors by the shape of the hand when the glove is worn. This phenomenon complicates the differentiation of trends in sensor values between different sign language gestures.

2) Baseline Correction: To mitigate the issue, we calculated the average value of the data at steady state for each sensor within each subject's data, on a per-subject basis. Then, we subtracted this subject-specific average from the corresponding sensor data for each subject. This process ensures that the baseline for each subject is adjusted individually, allowing for a more accurate comparison across all sign language data collected from the subjects. This adjustment was aimed at normalizing the steady-state values of the data to around zero. After applying Baseline Correction, the data were smoothed using a moving average filter.

3) Data Splitting: The processed data were then segmented according to individual sign language actions based on the videos recorded during the experiment. Subsequently, these segmented pieces of data were concatenated for each type of sign language gesture.

4) Data Normalization and Padding: Once the above processing was completed, we then applied Min-Max normalization separately to the data of each participant. In addition, due to the architectural requirements of LSTM and k-NN, it is necessary that all input data sequences are the same length. Therefore, we standardized the length of all data sequences to 100, and applied padding with -1.0 to any segments that were lacking in length.

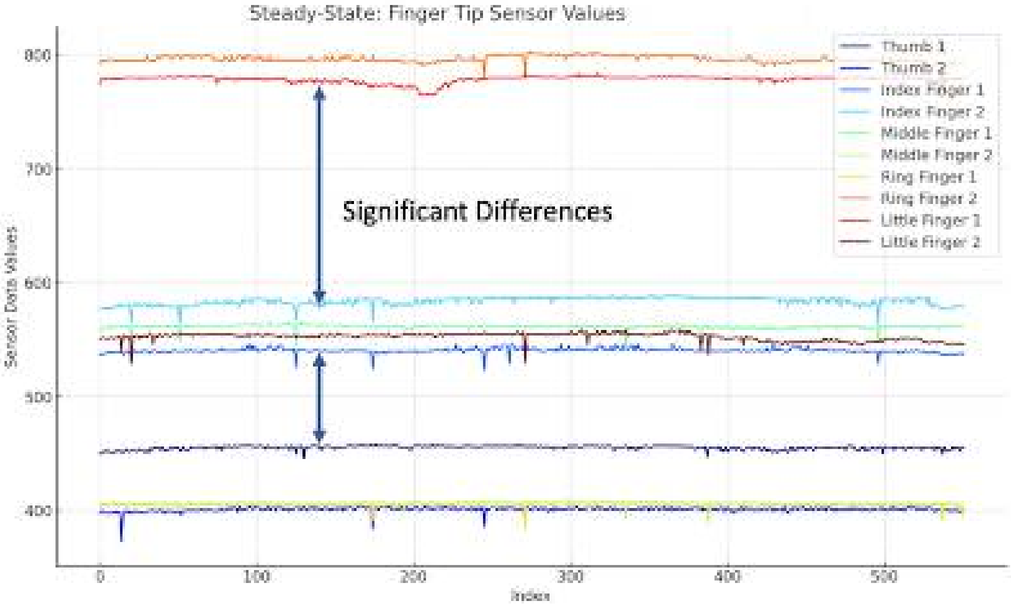


Fig. 9: The Issue with Data

V. RESULT

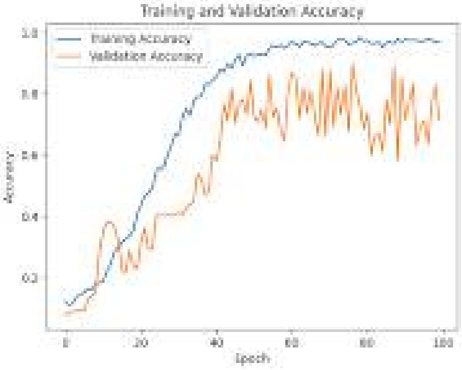


Fig. 10: Training and Validation Accuracy Trends Over Epochs for the LSTM model

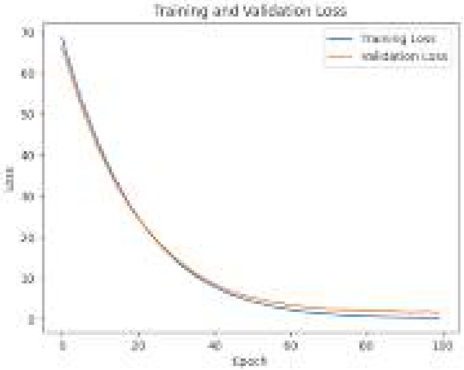


Fig. 11: Training and Validation Loss Trends Over Epochs for the LSTM model

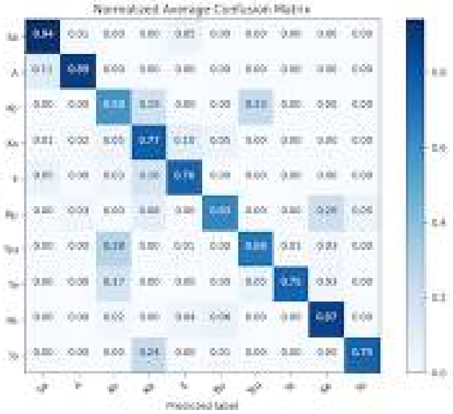


Fig. 12: Confusion Matrix of LSTM model

Metric	Value
Average Training Accuracy	0.91
Average Validation Accuracy	0.75
Average Training Loss	0.44
Average Validation Loss	1.03
Average Test Accuracy	0.76
Average Test Loss	1.02
Average Precision	0.71
Average Recall	0.72
Average F1 Score	0.76

TABLE I: LSTM: Training, Validation and Test Accuracy and Loss, Precision, Recall, F1 Score

As shown in Figure 11, we developed a streamlined LSTM model in Keras that includes a masking layer, a 32-unit LSTM layer with L2 regularization and a high dropout rate, Batch Normalization layer, and a 10-class softmax output layer, in order to improve the accuracy of sign language recognition while preventing overfitting. The model was trained for 100 epochs and compiled using Adam Optimizer with a learning rate set to 0.0005. Upon group 5-fold cross-validation, the results are presented in Table 1. From the confusion matrix presented in Figure 12, it is observed that ’Sa’ exhibited the highest positive rate with a value of 1.0, indicating perfect recognition. In contrast, ’Ko’ showed a significantly lower positive rate at 0.58, highlighting a disparity in model performance across different classes.

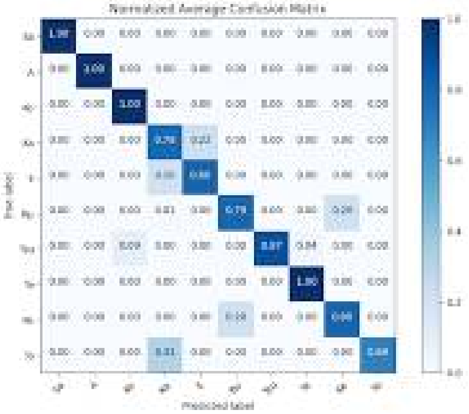


Fig. 13: Confusion Matrix of k-NN

Metric	Value
Average Accuracy	0.87
Average Precision	0.88
Average Recall	0.87
Average F1 Score	0.85

TABLE II: k-NN: Average values of Precision, Recall, and F1 Score

On the other hand, the results obtained using the k-NN algorithm indicated a relatively high average accuracy of 87%, as illustrated in Table 2. In the implementation, the ‘n_neighbors‘ parameter was set to 5, and the distance metric used was DTW. This configuration was chosen to effectively capture the temporal similarities in the time series data, enabling a detailed comparison of the temporal behavior of the dataset. Upon examining the Confusion Matrix, it was observed that four labels, specifically ’Sa’, ’A’, ’Ko’, and ’Te’, achieved a true positive rate of 1.0. In contrast, ’To’ showed a significantly lower positive rate at 0.69.

VI. Discussion

The LSTM model’s test accuracy is around 76%, indicating potential overfitting due to a significant gap between training and test accuracies. The fluctuation in validation accuracy suggests that the small validation set size may limit the model’s generalization capabilities, impacting its performance on unseen data. The k-NN model’s higher average accuracy of approximately 87% shows it to be a viable alternative, with a well-balanced performance across classes, despite a slightly lower F1 score hinting at possible precision-recall trade-offs.

Both models consistently achieved high accuracy for signs ’Sa’ and ’A’, with confusion mainly observed between ’Ka’ and ’To’. This misclassification is attributed to the subtle difference in the gestures, particularly the bending of the middle finger’s third joint, which was not sufficiently captured by the models. Despite distinct gestures for ’Ru’, ’Ke’, ’To’, and ’Ka’, similarities based on Dynamic Time Warping (DTW) distance calculations led to classification challenges. The limited dataset underscores the need for additional data, possibly through augmentation, to improve model understanding and accuracy.

VII. Conclusion & Future Work

In this study, we developed a novel tactile glove equipped with force-sensitive sensors and examined its effectiveness in sign language recognition. This glove can accurately capture the fine hand movements necessary for sign language communication. The evaluation results showed that both the LSTM and k-NN models achieved high accuracy, particularly demonstrating that our glove’s sensor placement is effective for sign language recognition. However, there is room for improvement in the F1 score.

Future research will focus on expanding the dataset to enhance the models’ generalization capabilities and validate their effectiveness. Specifically, we will consider introducing data augmentation techniques and including more diverse sign language gestures. Additionally, we aim to suppress overfitting in the LSTM model and fine-tune the k-NN model parameters to improve accuracy and F1 scores.

Ultimately, our goal is to develop a real-time sign language translation application. This application aims to bridge the communication gap between deaf and hearing individuals, contributing to the development of inclusive communication tools.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 22K12114 and NEDO Intensive Support for Young Promising Researchers Number 21502121-0 and JKA and its promotion funds from KEIRIN RACE.

References

[1] M. Ahmed, B. Zaidan, A. Zaidan, M. Salih, and M. Lakulu, “A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017,” *Sensors*, vol. 18, no. 7, p. 2208, 2018.

[2] D. Alosail, H. Aldolah, L. Alabdulwahab, A. Bashar, and M. Khan, “Smart glove for bi-lingual sign language recognition using machine learning,” in *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pp. 409–415, 2023.

[3] S. Senanayaka, R. Perera, W. Rankothge, S. Usgalhewa, H. Hettihewa, and P. Abeygunawardhana, “Continuous american sign language recognition using computer vision and deep learning technologies,” in *2022 IEEE Region 10 Symposium (TENSYMP)*, pp. 1–6, 2022.

[4] D. M. Madhiarasan and P. P. P. Roy, “A comprehensive review of sign language recognition: Different types, modalities, and datasets,” 2022.

[5] Q. Gao, L. Sun, C. Han, and J. Guo, “American sign language fingerspelling recognition using rgb-d and dfanet,” in *2022 China Automation Congress (CAC)*, pp. 3151–3156, 2022.

[6] B. Shi, X. Chen, Z. He, and R. Han, “Development of magnetic-sensor-based hand gesture recognition system for sign language,” in *2023 IEEE 6th International Electrical and Energy Conference (CIEEC)*, pp. 2302–2305, 2023.

[7] S. Sundaram, P. Kellnhofer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik, “Learning the signatures of the human grasp using a scalable tactile glove,” *Nature*, vol. 569, no. 7758, 2019.