





# A Few-Shot Approach to Sign Language Recognition: Can Learning One Language Enable Understanding of All?

Ragib Amin Nihal<sup>1</sup>(✉)  and Nawara Mahmood Broti<sup>2</sup> 

<sup>1</sup> Tokyo Institute of Technology, Tokyo, Japan

ragib@ra.sc.e.titech.ac.jp

<sup>2</sup> Meiji University, Tokyo, Japan

**Abstract.** Sign language is a unique form of communication in which hand or other body part gestures are used to express oneself. A large proportion of the world's population has speech and hearing impairments and communicates through sign language. Sign language, like verbal language, varies from country to country. Recent researches on automatic recognition focus on specific sign language of a country and require a large dataset. However, a prevalent issue arises when there is plenty of data available for some sign languages, while other sign languages suffer from data scarcity or non-existence of resources. To tackle this issue, our study presents a novel solution by proposing a few-shot learning approach for automatic sign language recognition. This approach involves training the model using data from a single sign language and then leveraging the acquired knowledge to recognize other sign languages, even when limited data is available for those languages. By bridging the gap between limited data availability and accurate recognition of new Sign Languages via employing this few-shot learning technique, our approach contributes to enhancing communication accessibility for the global sign language community. Our experimental results demonstrate promising performance, showcasing the potential of our model in overcoming the challenges of cross-lingual sign language recognition.

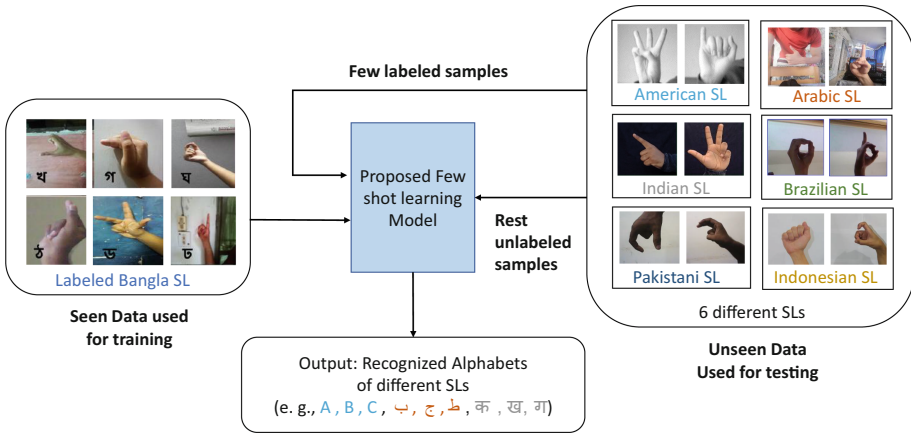
**Keywords:** Sign language · Few shot learning · automatic recognition · multiple sign language

## 1 Introduction

Sign language (SL) is a distinctive form of communication for hearing and speaking-impaired people that uses hand gestures and non-verbal indicators (such as facial expressions) to convey messages. Rather than using spoken communication and sound patterns, SL users communicate through signs in visual space. Based on a survey by the World Federation of the Deaf, there are more than 300 SLs used by 70 million deaf individuals worldwide [1]. Due to lack of knowledge, most of the hearing population do not understand this language,

which raises a barrier for SL users in society. Recently, deep learning-based automatic SL recognition systems are showing promise and may be able to help break down these barriers.

Every country has its unique SL, having its own set of gestures and meanings. However, only a handful of these SL data are publicly available; for the rest, data availability is extremely limited. One of the key reasons is that manually collecting and labeling enough SL samples is expensive and time-consuming. In addition, the data needs to be collected in the presence of SL experts for authenticity. As a result of data scarcity, less-popular SLs do not receive the necessary attention from researchers and remain unexplored.



**Fig. 1.** Overview of the proposed SL recognition system: Bangla SL alphabets are used as training data to train the proposed few-shot learning model; six other countries' SL images are used as test data. The model recognizes alphabets of different SLs from its knowledge from training data and a few labeled samples of other SLs. The few-shot learning approach enables effective recognition of SL alphabets across diverse linguistic contexts, enhancing cross-lingual SL recognition.

SL can be expressed with either motion or static hand gestures. In accordance with how an SL is expressed, the corresponding SL dataset can consist of images or videos. In many researches, traditional Convolutional neural network (CNN) and Recurrent neural network (RNN) based deep learning systems are proven to have extraordinary performance in SL recognition [3–5]. However, the majority of automatic SL recognition systems rely on large-scale data training. Without a large enough dataset, these techniques fail to perform. Recently, some innovative forms of learning is introduced that can learn from very few training data. Few-shot learning is one of them. Few-Shot Learning is a machine learning sub-field that is concerned about categorizing new data when we just have a few supervised training examples [2]. Lately, researchers are applying few-shot learning in automatic SL recognition systems. Wang et al. [6] proposed a metric-based few-shot learning model called Cornerstone Network for Chinese natural

SL recognition. Bohacek et al. [7] introduced an online text-to-video dictionary-based video dataset and a novel approach to training SL recognition models in a few-shot scenario. Shovkopliias et al. [8] utilized video data of hand gestures, as well as Electromyography signal from arm for few-shot learning based SL recognition. Ferreira, et al. [9] proposed a Contrastive Transformer-based model that can learn rich representations from body key points sequences. Hosseini et al. [10] introduced a one-shot learning approach for teaching new Iranian SL signs to a teacher assistant social robot. The performance of image-based few-shot learning in general SL recognition along with cross-lingual SL recognition is yet to be examined. This study tests the viability of learning many regions' static SLs using an image-based few-shot learning approach. As shown in Fig. 1, this work utilizes few-shot learning to learn different SL alphabets from few annotated samples and intends to provide a strong and adaptable tool for properly understanding the hearing and speaking communities. The contributions of the paper are summarized below:

1. **Few-shot Learning for Sign Language Recognition:** The research contributes to the application of few-shot learning techniques in the domain of SL recognition. By training a model on one SL dataset and leveraging the acquired knowledge to recognize signs from other SLs with limited data, the research addresses the data scarcity problem and opens up possibilities for more inclusive and effective SL recognition systems.
2. **Cross-Lingual Sign Language Recognition:** The research explores the potential for cross-lingual SL recognition by training a model on a single SL (Bangla SL) and evaluating its performance on multiple other SLs.
3. **Evaluation of Shot and Way Configurations:** The research conducts extensive experiments with varying shot and way configurations in the few-shot learning setup and selects DenseNet121 as the backbone architecture for feature extraction and Prototypical Networks as the few-shot classification method.

The rest of the paper is organized as follows: In Sect. 2, we discuss the selection and description of our datasets, providing a comprehensive overview of both the seen and unseen class datasets. Section 3 elaborates on the details of our methodology, including the problem statement, framework architecture, and experimental setup. We present the results of our experiments in Sect. 4, showcasing the performance of our model across different shot and way configurations. We then engage in a qualitative discussion in Sect. 5, highlighting the implications and limitations of our work, along with potential future directions. Finally, we conclude in Sect. 6 by summarizing our contributions and the significance of our approach in promoting inclusive communication through cross-lingual SL recognition.

## 2 Dataset Selection and Description

In the research, the selection of appropriate datasets plays a crucial role in training and evaluating the few-shot learning approach for SL recognition. The chosen

datasets should adequately represent the seen and unseen classes, capturing the diversity and variations present in different SLs.

1. Seen Class Dataset:

The seen class dataset- BdSL [11]- represents the Bangla SL. It is a large dataset of one-handed BdSL alphabet that contains 35,149 images. There are 37 different BdSL sign classes in this dataset, where each class possesses 950–1000 images. The images have versatile backgrounds and a broad range of light contrast, hand size, image scale, and skin tone of hand. The images are captured from more than 350 subjects and various angles. The image size is  $64 \times 64$  pixels. Due to the versatility of this dataset, it is selected as the primary source of labeled support samples for training the few-shot learning model.

2. Unseen Class Datasets:

The unseen class datasets, UdSL, encompasses six different SLs such as American SL (ASL) [12], Arabic SL (ArSL) [13], Brazilian SL (BrSL) [14], Indian SL (ISL) [15], Pakistani SL (PSL) [16], and Indonesian SL (InSL) [17]. These datasets have completely new SL images and new classes which are not present in the previously described seen class dataset. They represent the target SLs that the few-shot learning model aims to recognize using the knowledge acquired from the BdSL dataset. Each dataset is concisely described in Table 1.

**Table 1.** Unseen Dataset Description

Dataset	Total image number	Image size	Background	Color	Total class number
ASL	34627	$28 \times 28$	Homogenous	Gray scale	24
ArSL	5832	$416 \times 416$	Versatile	RGB	29
BrSL	4411	$200 \times 200$	Homogenous	RGB	15
ISL	42000	$128 \times 128$	Homogenous	RGB	35
PSL	1549	$640 \times 480$	Homogenous	RGB	38
InSL	520	$2000 \times 2000$	Homogenous	RGB	26

We selected the unseen class datasets considering the diversity of SLs, including variations in hand shapes, gestures, and cultural influences. This promotes the model’s ability to generalize to new SLs and improves its overall performance in recognizing signs from different countries.

### 3 Methodology

#### 3.1 Problem Statement

This research focuses on addressing the challenge of limited data availability for SL recognition systems. Here, *BdSL* represents the seen class, referring to

the Bangla SL dataset, and  $UdSL$  represents the unseen classes, comprising the  $ASL$ ,  $ArSL$ ,  $BrSL$ ,  $ISL$ ,  $PSL$ , and  $InSL$  datasets. The goal is to develop a few-shot learning approach that leverages knowledge acquired from the  $BdSL$  dataset to accurately recognize signs from the  $UdSL$  dataset.

Let  $X_s = (x_i, y_i)$  be the seen class dataset, where  $x_i$  represents the  $i$ th image sample and  $y_i$  denotes its corresponding class label from the  $BdSL$  dataset. Similarly, let  $X_u = (x_j, y_j)$  represent the unseen class dataset, where  $x_j$  is the  $j$ th image sample and  $y_j$  is its class label from the  $UdSL$  dataset.

The few-shot learning approach aims to train a model that can generalize knowledge from the seen class dataset ( $X_s$ ) to effectively recognize signs from the unseen class dataset ( $X_u$ ). A model is recurrently tested on a collection of  $N$ -way  $K$ -shot classification tasks, denoted as  $D_T = \{T_i\}$ , commonly known as episodes, using the standard few-shot learning procedure. To be more precise, each episode has a support set  $S_i$  and a query set  $Q_i$  split. The support set  $S_i$  has  $N$  distinct categories, each with  $K$  labeled samples, resulting in a total of  $N \times K$  examples for training. Each of the  $N$  categories in the query-set  $Q_i$  has  $Q$  unlabeled samples to categorize.

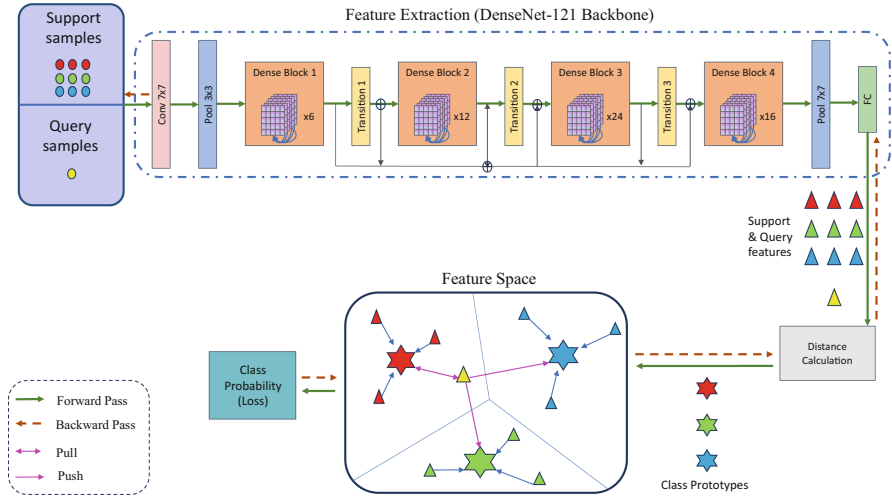
### 3.2 Overall Framework

In this research, we propose a framework that combines classical training using a CNN architecture with a few-shot learning based network. In the classical training phase, we train a powerful CNN architecture named DenseNet-121 [18]. This model was selected on the basis of our previous research [5, 19] that found that DenseNet extracts better features in SL recognition. We trained DenseNet-121 on the seen class dataset-  $BdSL$ . This process allows DenseNet-121 to learn discriminative features that capture rich representations of the input data. We optimize the network parameters using a cross-entropy loss across all training classes, which is a widely-used technique in traditional training setups.

The novelty of our research emerges in the second stage, where we introduce a few-shot learning framework. While the DenseNet-121 architecture itself is not a new contribution, the innovative aspect lies in its integration into a few-shot learning context. Specifically, we employ the prototypical network, a well-established approach in few-shot learning, to utilize the learned feature representations from the backbone network. This allows us to effectively recognize signs from unseen classes even when training data is limited.

Once DenseNet-121 is trained, it serves as the backbone network for the subsequent few-shot learning phase. Here, we employ the prototypical network [20], which is designed to handle few-shot learning tasks. In this phase, only a few labeled examples are taken from each unseen class from the unseen class dataset-  $UdSL$ . The prototypical network learns to compute class prototypes by averaging the feature embeddings of the support set, which consists of the labeled examples. These prototypes act as representatives of each class.

During testing, we extract feature embeddings of the query examples using the trained DenseNet-121 backbone. The prototypical network then calculates the similarity between the query embeddings and the class prototypes. Based on



**Fig. 2.** Proposed Architecture: It incorporates DenseNet-121 as the backbone for feature extraction. The extracted features are used for distance calculation. In the Prototypical Network, class prototypes are calculated by computing the mean of the support samples belonging to each class in the feature space. The distance between the query features and the class prototypes is calculated to determine the class probability for each query sample. The class probabilities are then obtained through the forward pass, followed by the backward pass for training. The model utilizes the pull and push mechanisms to enhance inter-class separability and intra-class compactness, respectively.

these similarities, the query examples are classified into their respective classes. During training, the network undergoes forward and backward passes, where the loss is computed based on the class probabilities and the ground truth labels. The loss is optimized through a combination of pull (attracting query features towards class prototypes of their respective classes) and push (encouraging separation between different class prototypes) operations. This approach allows the network to effectively generalize to unseen classes, even with limited labeled examples. Figure 2 illustrates the architecture of the proposed model.

By combining the strengths of classical training with DenseNet-121 and few-shot learning using the prototypical network, our proposed framework offers a robust and flexible solution for tackling few-shot learning challenges. It harnesses the discriminative power of DenseNet-121 for extracting informative features while leveraging the prototypical network to enable accurate classification with minimal labeled examples. This framework has the potential to benefit various applications in computer vision and pattern recognition domains where few-shot learning scenarios are encountered.

### 3.3 Experimental Setup

A series of experiments were carried out utilizing multiple shot and way combinations in order to assess the effectiveness of the few-shot learning model to SL recognition. Here, the number of training instances per class is represented by shot, whereas the number of classes in each training episode is represented by way. The studies evaluated our model’s capacity to generalize from a single seen SL (BdSL) to identify unseen SLs (ASL, ArSL, BrSL, ISL, PSL, and InSL) with few samples.

The experiments were conducted using a high-performance GPU RTX 2080. The implementation utilized Python programming language and popular deep learning libraries, including PyTorch and Torchvision. A batch size of 512 was used, and Parallel data loading processes were employed to enhance efficiency. The model parameters were optimized using the Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.1, momentum of 0.9, and weight decay of  $5 \times 10^{-4}$ . The training process spanned 200 epochs. A scheduler was incorporated to adjust the learning rate at specific milestones (150 and 180 epochs) using a gamma value of 0.1.

The model was trained using several shots and ways configurations, such as 1, 3, 5, 10, 20, 30, and 50 shots, as well as 5-way, 10-way, and all-way combinations. Note that, for each training episode, the model was fed with randomly selected shots from the seen class dataset.

### 3.4 Evaluation Metric

The trained model’s performance was evaluated using accuracy as the primary evaluation metric. Accuracy scores were calculated by comparing the model’s predictions with the ground truth labels of the unseen class datasets and were recorded for each shot and way configuration, providing a comprehensive assessment of the model’s recognition capabilities.

## 4 Results

The results of the few-shot learning experiments are summarized in the Table 2. The table presents the performance of different SLs (SL) in terms of accuracy for 1, 3, 5, 10, 20, 30, and 50 shots and 5-way, 10-way, and all-way combinations. The accuracy is normalized and reported on a range of 0 to 1, where 1 indicates 100% accuracy and 0 indicates 0% accuracy. Additionally, we plotted the results using line charts to provide a more intuitive representation of the model’s performance in Fig. 3. Note that, PSL and InSL datasets did not have enough samples to experiment with 20, 30, and 50 shots.

From the experimental results, we observed that the model’s performance improved as the number of shots increased, indicating that more training samples positively impacted its ability to generalize to unseen SLs. This trend was consistent across different numbers of ways (5, 10, and all) and for various SL

**Table 2.** Accuracy of the proposed model in different SL recognition (Measured on a Scale of 0 to 1.)

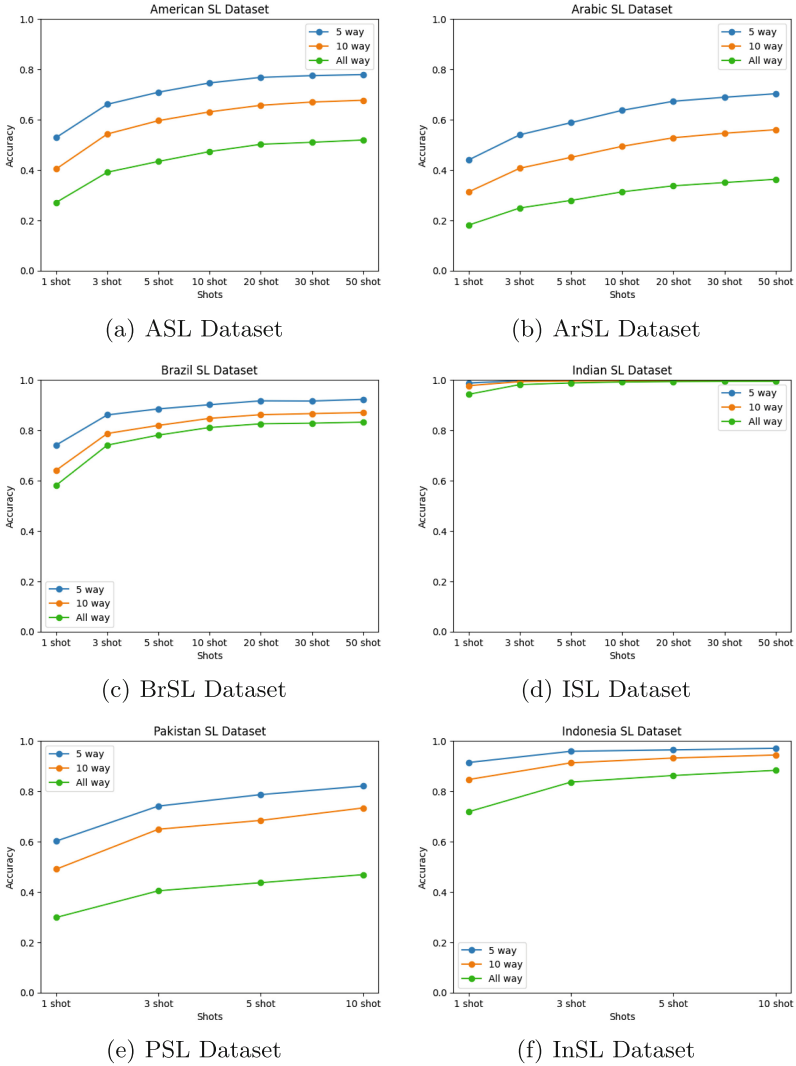
SL Name	Way	1-shot	3-shot	5-shot	10-shot	20-shot	30-shot	50-shot
ASL	5 way	0.529	0.661	0.709	0.746	0.768	0.775	0.779
	10 way	0.404	0.543	0.596	0.631	0.657	0.67	0.677
	All way	0.271	0.391	0.434	0.473	0.502	0.51	0.519
ArSL	5 way	0.44	0.54	0.588	0.637	0.673	0.689	0.703
	10 way	0.313	0.407	0.45	0.494	0.528	0.546	0.56
	All way	0.181	0.249	0.279	0.313	0.337	0.35	0.363
BrSL	5 way	0.741	0.861	0.884	0.902	0.917	0.916	0.923
	10 way	0.641	0.787	0.819	0.847	0.862	0.866	0.870
	All way	0.581	0.741	0.780	0.811	0.826	0.828	0.832
ISL	5 way	0.988	0.997	0.998	0.998	0.998	0.999	0.999
	10 way	0.977	0.993	0.996	0.997	0.997	0.997	0.998
	All way	0.943	0.982	0.988	0.992	0.993	0.994	0.994
PSL*	5 way	0.602	0.741	0.786	0.820	–	–	–
	10 way	0.490	0.649	0.684	0.7337	–	–	–
	All way	0.298	0.404	0.436	0.468	–	–	–
InSL*	5 way	0.914	0.959	0.964	0.970	–	–	–
	10 way	0.847	0.913	0.932	0.944	–	–	–
	All way	0.719	0.837	0.863	0.883	–	–	–

\*PSL and InSL datasets did not have enough samples to experiment with 20, 30, and 50 shots.

datasets. Notably, the model achieved higher accuracy in the 5-way scenario compared to the 10-way and all-way scenarios, suggesting that distinguishing between fewer classes was relatively easier for the model.

The model performed the best in classifying ISL, with a 5-way, 10-shot accuracy of 0.998 and an all-way, 10-shot accuracy of 0.992. The second and third best performance was achieved in InSL and BrSL classification with 5-way, 10-shot accuracy of 0.97 and 0.902 respectively. In PSL classification, the model achieved an accuracy of 0.820 in the 5-way, 10-shot scenario. In the case of the ASL and ArSL datasets, the model achieved an accuracy of 0.746 and 0.637 respectively in the 5-way, 10-shot scenario. These results demonstrate the model’s ability to recognize SLs even with limited training data, indicating the effectiveness of the few-shot learning approach.



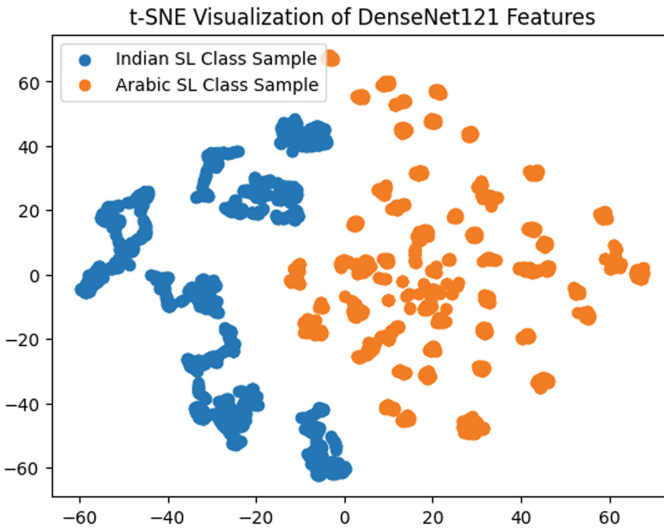


**Fig. 3.** Performance of the proposed model in different SL datasets

## 5 Discussion

In this work, we attempted to develop a machine learning model that utilizes few-shot learning in new SL classification from very few training data. From the performance of our proposed model, we have some observations reported as follows.

**1. Variations in Datasets:** Comparing the performance across different SLs, we observed variations in accuracy. For example, the ISL dataset exhibited consistently high accuracy across different shot and way scenarios, with an accuracy of 0.999 in the 5-way, 50-shot scenario. On the other hand, the ArSL dataset showed relatively lower performance, with a 5-way, 50-shot accuracy of 0.703. This could be attributed to the complexity or diversity of the signs within that particular language. To gain a better understanding of the differences between these two SLs, we performed t-distributed stochastic neighbor embedding (t-SNE) analysis. We plotted the t-SNE visualization of the features extracted from a class in both the ISL and ArSL datasets (Fig. 4). The t-SNE plot provides a visual representation of the similarity and proximity of the feature vectors.



**Fig. 4.** t-SNE Visualization of Indian SL and Arabic SL features. ISL class features exhibit closer proximity and similarity to each other, indicating a higher degree of intra-class consistency. On the other hand, ArSL class features appear more diverse and spread out in the feature space, suggesting greater inter-class variation.

From the t-SNE plot, it is evident that the features of ISL class samples are more similar and closely clustered with each other, indicating a higher degree of consistency and similarity in the visual patterns of the signs. This cohesion in the feature space can contribute to the model's ability to better discriminate between different ISL signs. On the other hand, the t-SNE plot of the ArSL class features shows more diversity and dispersion among the samples. This suggests that the ArSL signs exhibit greater variability and visual dissimilarity, which can pose challenges for the model in accurately differentiating between them. The scattered nature of the ArSL features indicates the presence of distinct subgroups or variations within the ArSL dataset, making it more difficult for the model to generalize effectively.

**2. Convergence Behavior:** Another observation from our research is the convergence behavior of the model as we increase the number of shots. We noticed that as the number of shots increased, the model’s performance tended to converge to a certain level of accuracy. This convergence behavior indicates that there is a threshold, beyond which providing additional training samples does not significantly improve the model’s ability to generalize to unseen SLs.

For example, in the 5-way scenario, we observed that the accuracy of the model gradually improved as the number of shots increased from 1 to 20. However, beyond a certain point, such as 20 or 30 shots, the improvement in accuracy became marginal, and the model’s performance stabilized. This suggests that once the model has sufficient exposure to a variety of samples for each class, further increasing the number of shots does not lead to substantial gains in accuracy.

**3. Cross-Lingual Sign Language Recognition:** The recognition of cross-lingual SLs is a difficult and crucial task in the field of computer vision and machine learning. Individuals with speech and hearing impairments stand to gain significantly from the ability to understand SLs in other languages, as well as communication and inclusivity on a global scale.

The outcomes of our research show how useful this method is for enabling cross-lingual SL recognition. We achieved promising accuracy across various shot and way scenarios by training the model on a single SL dataset and using the learned knowledge to recognize other SLs, even with limited data. However, further research and experimentation are necessary to explore additional techniques and architectures that can improve the accuracy and robustness of the model in diverse SL recognition scenarios.

**4. Limitations:** Video and action recognition methods must frequently be used for real-world SL recognition. Despite the fact that our study was primarily concerned with image-based recognition using still frames, moving the strategy to video-based recognition would be beneficial for capturing the temporal dynamics and movement patterns that are unique to SL.

The performance of our model heavily relies on the available SL datasets. We used a limited number of SL datasets for testing, which may not fully represent the vast diversity of SLs worldwide. Incorporating a wider range of SLs and dialects would enhance the generalizability and robustness of the model.

## 6 Conclusion

In this study, we proposed a few-shot learning approach for automatic sign language recognition to address the problem of data scarcity and limited resources. Our proposed approach combines DenseNet-121 architecture and prototypical network to leverage the information from small datasets in accurate recognition of different countries of the world. We trained our proposed model with a large

seen dataset (BdSL) and evaluated the model with six different unseen datasets from six different countries. Our approach offers several strengths and contributions that hold significance for both the research community and practical applications. We have showcased the adaptability and versatility of few-shot learning by training on a single sign language dataset and effectively recognizing signs from various other languages. This not only addresses the data scarcity challenge but also paves the way for more inclusive communication for individuals with speech and hearing impairments. Through extensive experimentation, we have demonstrated the effectiveness of our model in recognizing sign languages with limited data availability. Our model achieved impressive performance across various shot and way scenarios, with accuracy ranging from 18.1% to 99.9% on different datasets. These results highlight the potential of our approach in overcoming the data scarcity problem and enabling accurate sign language recognition across multiple languages. Moreover, we observed a converging behavior of the model and noticed that the model performance converges around 20 shots. The findings of our research indicate the significance of this approach for achieving cross-lingual sign language recognition. However, it is important to note that there are still areas for improvement. The proposed system is only verified on alphabet-based static SLs where there are many dynamic word or sentence level SLs that still require being brought under investigation. Future research could focus on developing strategies to handle the nuances of complex signs and gestures, potentially through the integration of temporal information from video data. Additionally, the computational efficiency of our approach could be optimized to facilitate real-time applications, ensuring its practical usability in real-world scenarios. Exploring other advanced deep learning architectures and expanding the range of datasets may lead to better performance and generalization across diverse sign languages. Furthermore, collaborative efforts between experts in SL linguistics and machine learning could lead to the development of more refined and culturally sensitive models for diverse sign languages. We hope that this research will help other researchers learn more about sign language recognition, benefiting hearing and speech-disabled people from all over the world. Further development and exploration of our proposed few-shot learning approach hold promise for improving cross-lingual sign language recognition systems to contribute to a wider sign language community.

## References

1. World Federation of the Deaf. <https://wfdeaf.org/>. Accessed 21 May 2023
2. Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M.: Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv. (csur)* **53**(3), 1–34 (2020). <https://doi.org/10.1145/3386252>
3. Oyedotun, O.K., Khashman, A.: Deep learning in vision-based static hand gesture recognition. *Neural Comput. Appl.* **28**(12), 3941–3951 (2016). <https://doi.org/10.1007/s00521-016-2294-8>
4. Bantupalli, K., Xie, Y.: American sign language recognition using deep learning and computer vision. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 4896–4899 (2018). <https://doi.org/10.1109/BigData.2018.8622141>

5. Nihal, R.A., Broti, N.M., Deowan, S.A., Rahman, S.: Design and development of a humanoid robot for sign language interpretation. *SN Comput. Sci.* **2**(3), 1–17 (2021). <https://doi.org/10.1007/s42979-021-00627-3>
6. Wang, F., et al.: Cornerstone network with feature extractor: a metric-based few-shot model for Chinese natural sign language. *Appl. Intell.* **51**(10), 7139–7150 (2021). <https://doi.org/10.1007/s10489-020-02170-9>
7. Bohacek, M., Hruz, M.: Learning from what is already out there: few-shot sign language recognition with online dictionaries. In: 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), pp. 1–6. IEEE (2023). <https://doi.org/10.1109/FG57933.2023.10042544>
8. Shovkoplias, G.F., et al.: Improving sign language processing via few-shot machine learning. *Sci. Tech. J. Inf. Technol. Mech. Opt.* **22**(3), 559–566 (2022)
9. Ferreira, S., Costa, E., Dahia, M., Rocha, J.: A transformer-based contrastive learning approach for few-shot sign language recognition. *arXiv preprint arXiv:2204.02803* (2022)
10. Hosseini, S. R., Taheri, A., Alemi, M., Meghdari, A.: One-shot learning from demonstration approach toward a reciprocal sign language-based HRI. *Int. J. Social Rob.* 1–13 (2021). <https://doi.org/10.1007/s12369-021-00818-1>
11. Nihal, R. A., Broti, N. M.: BdSL-MNIST, Mendeley Data, V. 1 (2023). <https://doi.org/10.17632/6f2wm5p3vf.1>
12. Sign Language MNIST. <https://www.kaggle.com/datasets/datamunge/sign-language-mnist>. Accessed 25 June 2023
13. Arabic Sign Language ArSL dataset. <https://www.kaggle.com/datasets/sabribeImadoui/arabic-sign-language-unaugmented-dataset>. Accessed 25 June 2023
14. Passos, B.T., Fernandes, A.M.R., Comunello, E.: Brazilian Sign Language Alphabet, Mendeley Data, V. 5 (2020). <https://doi.org/10.17632/k4gs3bmx5k.5>
15. Indian Sign Language Dataset. <https://www.kaggle.com/datasets/vaishnaviasonawane/indian-sign-language-dataset>. Accessed 25 June 2023
16. Pakistan Sign Language. <https://www.kaggle.com/datasets/hasaniqbal777/pakistan-sign-language>. Accessed 25 June 2023
17. Mursita, R.A.: Respon tunarungu terhadap penggunaan sistem bahasa isyarat indonesa (sibi) dan bahasa isyarat indonesia (bisindo) dalam komunikasi. *Inklusi* **2**(2), 221–232 (2015). <https://doi.org/10.14421/ijds.2202>
18. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
19. Nihal, R.A., Rahman, S., Broti, N.M., Deowan, S.A.: Bangla sign alphabet recognition with zero-shot and transfer learning. *Pattern Recogn. Lett.* **150**, 84–93 (2021). <https://doi.org/10.1016/j.patrec.2021.06.020>
20. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. *Adv. Neural. Inf. Process. Syst.* **30**, 1–11 (2017)