Sign Language Recognition Using Artificial Intelligence

Willis Gotama
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
willis.gotama@binus.ac.id

Reinardus Ariel Joan Anandika
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
reinardus.anandika@binus.ac.id

Alif Tri Handoyo
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
alif.handoyo@binus.ac.id

Edy Irwansyah
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
edirwan@binus.ac.id

***Abstract*— Hearing loss is a problem that can occur due to many factors. People with hearing loss find it difficult to communicate with other people because of communication limitations. The presence of this problem led to sign language being created to help communication between deaf people or people with hearing impairments. However, many people are still unfamiliar with sign language and still find it difficult to communicate with deaf people. As time progresses, many technologies help make it easier for deaf people to communicate with people without hearing impairments. Previous researchers have used many methods to create technology that can help translate sign language into words in both spoken and text form. In this research, we discuss sign language recognition using the CNN model. This research aims to develop and test a simple CNN model for recognizing American Sign Language (ASL). This model was tested using 3 different data sets with different data. The research results show different levels of accuracy between data sets but have the highest accuracy level of 100%. However, this research has limitations in its implementation where the model was not compared with other models and the model was not tested with original data in the field. Therefore, further research is needed to address limitations and ensure the model works well with existing data in the field.**

Keywords—CNN, ASL, American Sign Language

I. INTRODUCTION

Hearing loss is a very difficult problem. Hearing loss can be caused by various factors such as age, genetic disorders, and the environment. Environments that have a high level of distraction have a greater potential for hearing loss and trauma [1]. For example, constant exposure to loud sounds in the workplace or noisy urban environments can permanently damage hearing. Social communication will be difficult if the person you are talking to has hearing loss because normal speech sounds jumbled and confusing. This not only impacts daily interactions, but can also impact the mental health of individuals who experience difficulty communicating. A solution that can be used to continue communicating is to use sign language. Sign language is a communication medium that uses body movements and facial expressions in human communication.

According to the British Deaf Association, as many as 151,000 people in the UK communicate using iconic sign language[2]. In the United States, more than 500,000 people use American Sign Language (ASL) as their primary language. There is no universal sign language, but there are many types of sign language and almost every country has a sign language used such as American Sign Language (ASL), Indian Sign Language (ISL), and Devanagari Sign Language (DSL) [3]. The use of sign language is not only important for people with hearing disabilities, but also helps strengthen inclusive communication in society at large. By learning and using sign language, society can better interact with deaf people, encourage inclusion, and reduce the social stigma that may exist towards hearing loss.

However, the application of sign language in everyday life still faces many challenges. One of the main challenges is the low understanding and skills of sign language among the general population, which often causes difficulties in communication between deaf individuals and those who are not trained in sign language, let alone communication between deaf and dumb people and hearing people[4]. In addition, the infrastructure and technology supporting sign language translation are not yet fully developed in many places, so access to information and services for the deaf community is still limited. For example, in many countries, sign language interpreting services in hospitals, courts and other public institutions remain inadequate, which can hinder the basic rights and needs of deaf people.

Therefore, developing a sign language translator application is very important. This application can not only help individuals with hearing loss communicate more effectively, but also support social inclusion efforts more broadly. However, other problems arise because there are many techniques that can be used to create sign language translator applications. Each technique has its own advantages and disadvantages, and choosing the right technique is critical to achieving accurate and reliable results. Through this article, we will conduct an experiment to identify the accuracy of American Sign Language (ASL) recognition. We hope that the results of this research can make a significant contribution to the progress of sign language translation technology and improve the quality of communication for people with hearing disabilities. With the right technology, we can create a more inclusive society and provide greater opportunities for everyone, regardless of their hearing ability.

II. Related works

Sign language is a topic that has been studied for a long time. Many methods are used to help translate sign language so that it is easy to understand. The methods used in sign language translation are combined with sensors such as cameras and gloves. The data received by the sensor will be processed and processed using methods that have been created to be translated and produce the desired output.

Karhunen-Loeve Transforms is a transformation technique used by Singha and Das to reduce dimensions and identify important features in data [5]. In their approach[5], hand images are processed using skin filtration techniques to help identify important parts of the hand and match them using Euclidean Distance to match the results with a database.

In this journal[6], researchers used the k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) approaches in research to distinguish hand shapes from the background by detecting skin color using RGB and YCbCr color spaces as well as level intensity thresholds. gray to produce accurate classification results.

Abdou in his journal [7] uses The Recurrent Neural Network (RNN) combined with a Graphical Processing Unit (GPU) to create a neural network-based system that is evaluated using objective and subjective measures to reduce error rates and make it possible to capture hand movements and facial expressions accurately.

In overcoming social problems for deaf people, a previous research journal[8] created a sign language translation system using Microsoft Kinect, convolutional neural networks (CNNs) and GPU acceleration. In research they conducted using CNN[8], there were 20 Italian sign languages that were recognized with a very high level of accuracy and were able to adapt to the user and the surrounding environment as proven by a validation accuracy of 91,7%.

RNN is a model in deep learning that can process data. In the research journal [9] the problem of sign language translation is discussed and developed using Recurrent Neural Network (RNN). The mechanism used is Long Short-Term Memory (LSTM) and combined with Gated Recurrent Units (GRU) to optimize system performance in detecting sign language from videos and translating them into English by dividing the video into small frames and identifying the beginning and end . every move.

In a previous journal [10] research was conducted using a 3D recurrent convolutional neural network (3DRCNN) model to recognize American Sign Language. The model also uses a fully connected recurrent neural network (FC-RNN) to capture temporal information. Evaluation of this model was carried out by collecting new ASL data with an accuracy of 69.2%.

Research that has been carried out by other researchers [11] uses a transfer learning architectural model in developing a translation system for American Sign Language (ASL) into English and vice versa. The main purpose of using transfer learning in this research is because there is already a model that has been trained previously and has similar information and data. The model used in this research is ResNet50 which has previously been trained on image data and has good accuracy. Using this model also makes it easier for researchers when they want to increase the number of network layers without reducing the level of accuracy.

III. Metodology

Given a dataset in the form of sign language images, where the images are taken from the American sign language(ASL). The datasets we use are ASL Alphabet, hand-sign-images, and ASL(American Sign Language) Alphabet dataset. The dataset that has been obtained is then labeled correctly according to the instructions. Then from these images we carry out training using the CNN model to carry out image classification. For each image and word, train and test are carried out for each dataset.

The architecture of the CNN was designed to efficiently recognize patterns and features within the sign language images. The labeled datasets were split into training, validation, and test sets with respective ratios. The training set was used to train the CNN model, while the validation set was used to tune the model parameters and prevent overfitting. Finally, the test set was employed to evaluate the model's performance on unseen data.

3.1 Dataset

The dataset studied was taken from American Sign Language (ASL). This dataset consists of three different datasets where each dataset has a different number of images. Each dataset has the alphabet (A-Z) displayed in hand-drawn form.

The ASL Alphabet contains 28 images in the form of the alphabet from A-Z and additions in the form of delete, space, and nothing. Meanwhile, the ASL (American Sign Language) Alphabet dataset consists of 2 files in the form of train and test, where each file contains 29 images. Additionally, the third dataset, hand-sign-images, was also used in this research. This dataset consists of various hand images showing the alphabet letters in ASL.

From the collected dataset, training and testing were carried out for each dataset. Training was conducted to train the CNN model to recognize each hand sign, while testing was conducted to evaluate the performance of the trained model. From each training and testing conducted on each dataset, the total results will be calculated to determine the accuracy and effectiveness of the model.

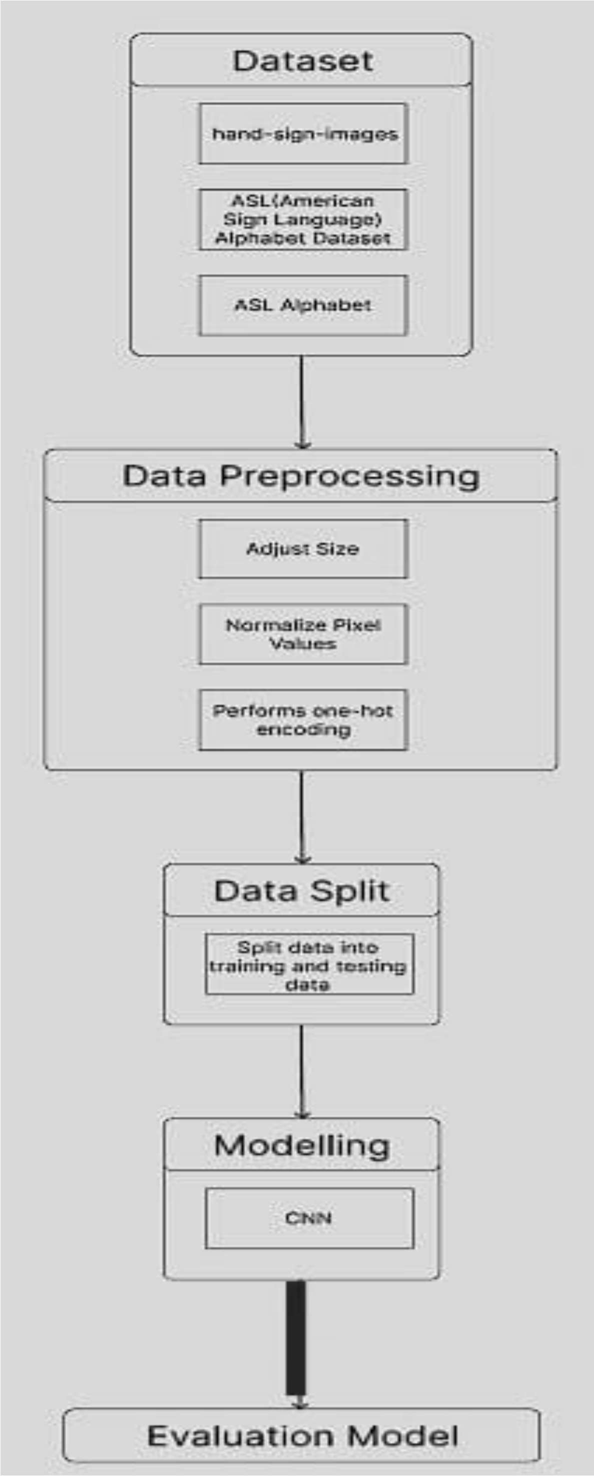
The use of these diverse datasets aims to ensure that the developed model has good generalization and can recognize various variations of hand signs in ASL. Thus, the resulting model is expected to be effectively used in real applications to assist communication for ASL users.

Details of every dataset used in this research, which are ASL Alphabet, hand-sign-images, and ASL Alphabet dataset, can be seen in Table 1. This table provides detailed information about the number of images, the type of images, and the data distribution for each dataset.

Dataset	Train	Test	Total
ASL Alphabet	70,200	7,800	78,000
hand-sign-images	24,709	2,746	27,455
ASL(American Sign	185,523	20,614	206137

Language) Alphabet Dataset			

Table 1. Dataset Table



3.2 Preprocessing

In this stage, preprocessing is carried out in the development of the CNN model that we aim to create. This preprocessing stage is conducted to ensure that the input data is in an optimal format for model training.

The first step undertaken is data collection, where in the previous stage, we selected a dataset consisting of hand images demonstrating various sign language gestures. We selected three datasets, each containing different amounts of data.

Next, after the data was collected, we performed data cleaning. In this data cleaning process, we removed blurry, irrelevant, or low-quality images. This process ensures that the model is trained with representative and high-quality data.

Then, we normalized the images to ensure that the pixel values were within a consistent range. Each image pixel was normalized by dividing the pixel value by 255. This helps to accelerate model convergence during training.

The images in the dataset were resized to uniform dimensions. This resizing is necessary for the CNN model as it requires input with consistent sizes. Additionally, smaller image sizes can reduce computational load.

After that, the images were converted to grayscale to reduce model complexity. By eliminating color information, the model can focus on the patterns and shapes of hand gestures, which is crucial information in sign language recognition.

The dataset was divided into three subsets: training set, validation set, and test set with ratios of 70%, 15%, and 15%, respectively. This division is important to ensure the model can be effectively tested and validated.

3.3 Modeling

At this stage, we developed a CNN (Convolutional Neural Network) model to recognize sign language. During the modeling phase, we developed this recognition system, where we designed, trained, and tested the model to achieve optimal performance. We conducted training on the dataset, which consists of the ASL Alphabet, hand-sign-images, and ASL (American Sign Language) Alphabet Dataset. For these three datasets, we performed training using a batch size of 64 and epochs size of 5.

IV. Result and Discussion

In this section, evaluation results from experiments using 3 different datasets are given: ASL Alphabet, Hand-sign-images, and ASL Alphabet Dataset. Experiments were carried out using a simple CNN model. This experiment was carried out to detect the level of accuracy produced by the CNN model in sign language recognition.

4.1 Model Performance

From experiments using the ASL alphabet with results as in table 2 there was an increase from the first Epoch with Train Accuracy of 80% and reaching the highest Training Accuracy of 98%. Validation Accuracy also increased and reached the highest accuracy of 98% in the fourth epoch but there was a decline in the fifth epoch. The model works well in learning the training data as indicated by a decrease in Train Loss and Val Loss. Overall, the model performs well in learning this dataset with a Test Accuracy of 96.73% and a Test Loss of 8.92%.

Epoch	1	2	3	4	5
Train Loss	0.6306	0.0778	0.0467	0.0347	0.0317
Train Accuracy	0.8044	0.9763	0.9858	0.9891	0.9897
Val Loss	0.1519	0.1054	0.0618	0.0349	0.0948
Val Accuracy	0.9550	0.9725	0.9798	0.9887	0.9637
Train time	2425.2133				
Test accuracy	0.9673				
Test loss	0.0892				
Test time	15.5454				

Table 2. ASL Alphabet dataset experimental results

From experiments using the Hand-sign-images dataset with results as in table 3, there was not much change in Train Accuracy. In the first Epoch, Train Accuracy was 89.48% and Validation Accuracy was 99.96%, while Training Accuracy and Validation Accuracy in other Epochs reached 100%. The Train Loss and Val Loss produced by this model using the Hand-sign-images dataset are very low, almost close to 0. Overall, the model works very well with a Test Accuracy of 100% and a Test Loss of 0.05% which shows that there is a possibility in the dataset this is easier for the model to learn.

Epoch	1	2	3	4	5
Train Loss	0.3820	0.0078	0.0025	0.0014	0.0011
Train Accuracy	0.8948	1.0000	1.0000	1.0000	1.0000
Val Loss	0.1508	0.0024	5.1177e-04	3.4509e-04	5.8793e-04
Val Accuracy	0.9996	1.0000	1.0000	1.0000	1.0000
Train time	905.3007				
Test accuracy	1.0				
Test loss	0.0005				
Test time	10.3402				

Table 3. Hand-sign-images dataset experimental results

Epoch	1	2	3	4	5
Train Loss	0.6046	0.1388	0.0898	0.0706	0.0604
Train Accuracy	0.8189	0.9563	0.9717	0.9769	0.9802
Val Loss	0.1196	0.0641	0.0571	0.0431	0.0323
Val Accuracy	0.9650	0.9801	0.9826	0.9873	0.9895
Train time	6567.5085				
Test accuracy	0.9897				
Test loss	0.0313				
Test time	83.3922				

Table 4. ASL(American Sign Language) Alphabet Dataset experimental results

The CNN model also works very well using the ASL (American Sign Language) Alphabet Dataset as shown in table 4. In the first Epoch, Train Accuracy was 81.89% with Train Loss of 60.46%. In the next Epoch, Train Accuracy increased to 95.63% and Train Loss decreased to 13.88%. Train Accuracy and Train Loss reached the highest and lowest values in the fifth epoch of 98.02% and 6.04%. Val Accuracy in this dataset has touched 96.5% in the first epoch and the highest was 98.95% in the fifth epoch, likewise Val Loss reached its lowest level in the fifth epoch with 3.23%. Overall the model can perform well learning this dataset with a Test Accuracy of 90.97% and a Test Loss of 3.13%

Of the three datasets used in training, the ASL (American Sign Language) Alphabet Dataset has the longest Train Time because it contains more than 185,000 data, while the Hand-sign-images dataset has the shortest Train Time because it only has less than 25,000 Train data. This shows that the size of the dataset affects the duration of training.

V. Conclusion

Based on the experimental data that has been carried out, the simple CNN model is able to recognize sign language very well and shows excellent performance using these three data sets. The best accuracy was obtained when the model was tested using the Hand-sign-images data set with an accuracy level of 100%. In this way, this experiment shows the effectiveness of the Simple CNN model in sign language recognition. Unfortunately, this research has limitations that need to be considered, such as the model tested cannot be compared with other models due to time constraints so it cannot be concluded whether the model tested is better than other models. Limitations in accessing real data also mean that this research cannot be verified with real conditions in the environment.

References

[1] Nadol, J. B. (1993). Hearing Loss. New England Journal of Medicine, 329(15), 1092–1102. https://doi.org/10.1056/NEJM199310073291507

[2] Jalal, M. A., Chen, R., Moore, R. K., & Mihaylova, L. (2018). American Sign Language Posture Understanding with Deep Neural Networks. 2018 21st International Conference on Information Fusion (FUSION), 573–579. https://doi.org/10.23919/ICIF.2018.8455725

[3] Nikhil Kulkarni, Shivali Mate, Atharva Kulkarni, & Shailaja Jadhav. (2022). Sign Language Recognition. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 200–204. https://doi.org/10.32628/cseit228234

[4] Ibrahim, N., Zayed, H., & Selim, M. (2019). Advances, Challenges, and Opportunities in Continuous Sign Language Recognition. Journal of Engineering and Applied Sciences, 15, 1205–1227. https://doi.org/10.36478/jeasci.2020.1205.1227

[5] Singha, J., & Das, K. (2013). Hand Gesture Recognition Based on Karhunen-Loeve Transform. ArXiv, abs/1306.2599. https://api.semanticscholar.org/CorpusID:15113325

[6] Sharma, R., Nemani, Y., Kumar, S., Kane, L., & Khanna, P. (2013). Recognition of Single Handed Sign Language Gestures using Contour Tracing Descriptor. In Lecture Notes in Engineering and Computer Science (Vol. 2). https://api.semanticscholar.org/CorpusID:17294056

[7] Abdou, M. A.-E. (2018). An Enhanced Training- Based Arabic Sign Language Virtual Interpreter Using Parallel Recurrent Neural

Networks. J. Comput. Sci., 14, 228–237. https://api.semanticscholar.org/CorpusID:4601568

[8] Pigou, L., Dieleman, S., Kindermans, P.-J., & Schrauwen, B. (2015). Sign Language Recognition Using Convolutional Neural Networks. In L. Agapito, M. M. Bronstein, & C. Rother (Eds.), Computer Vision - ECCV 2014 Workshops (pp. 572–578). Springer International Publishing.

[9] Kothadiya, D. R., Bhatt, C. M., Sapariya, K., Patel, K. R., Gil-González, A. B., & Corchado, J. M. (2022). Deepsign: Sign Language Detection and Recognition Using Deep Learning. Electronics. https://api.semanticscholar.org/CorpusID:249335784

[10] Ye, Y., Tian, Y., Huenerfauth, M., & Liu, J. (2018). Recognizing american sign language gestures from within continuous videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2064–2073. http://openaccess.thecvf.com/content_cvpr_2018_workshops/w41/html/Ye_Recognizing_American_Sign_CVPR_2018_paper.html

[11] Avina, V. D., Amiruzzaman, M., Amiruzzaman, S., Ngo, L. B., & Dewan, M. A. A. (2023). An AI-Based Framework for Translating American Sign Language to English and Vice Versa. Information, 14(10). https://doi.org/10.3390/info14100569