Computer Methods and Programs in Biomedicine Update 2 (2022) 100048

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine Update

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine-update

DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals

Ahmed KASAPBAŞI [a], Ahmed Eltayeb AHMED ELBUSHRA [a], Omar AL-HARDANEE [a], Arif YILMAZ [b,*]

[a] Department of Electrical and Electronics Engineering, Graduate School of Natural and Applied Science, Ankara Yildirim Beyazit University, Ankara, Turkey
[b] Maastricht University, Institute of Data Science, Maastricht, Netherlands

ARTICLE INFO

Keywords:
Deep learning
Convolutional neural network (CNN)
Sign language recognition (SLR)
OpenCV
New Dataset

ABSTRACT

Background: Sign language is an essential means of communication for hearing-impaired individuals.
Objective: We aimed to develop an American sign language recognition dataset and use it in the deep learning model which depends on neural networks to interpret gestures of sign language and hand poses to natural language.
Methods: In this study, we developed a dataset and a Convolutional Neural Network-based sign language interface system to interpret gestures of sign language and hand poses to natural language. The neural network developed in this study is a Convolutional Neural Network (CNN) which enhances the predictability of the American Sign Language alphabet (ASLA). This research establishes a new dataset of the American Sign Language alphabet which takes into consideration various conditions such as lighting and distance.
Results: The dataset created in this study is a new addition in the field of sign language recognition (SLR). This dataset may be used to develop SLR systems. Furthermore, our research compares the results of our dataset with two different datasets from other studies. The other datasets have invariant scene conditions, but our suggested CNN model demonstrated high accuracy for all the tested datasets. Despite the different conditions and volume of the new dataset, it achieved 99.38% accuracy with excellent prediction and small loss (0.0250).
Conclusions: The proposed system may be considered a promising solution in medical applications that use deep learning with superior accuracy. Moreover, our dataset was created under variable conditions which increases the number of contributions, comparisons, results and conclusions in the field of SLR and may enhance such systems.

Introduction

More than 5% of the world's population is affected by hearing impairment. To overcome the challenges faced by these individuals, various sign languages have been developed as an easy and efficient means of communication. Sign language depends on signs and gestures which give meaning to something during communication [1]. Researchers are actively investigating methods to develop sign language recognition systems, but they face many challenges during the implementation of such systems which include recognition of hand poses and gestures. Furthermore, some signs have similar appearances which add to the complexity in creating recognition systems [3,4]. This paper focuses on the sign language alphabet recognition system because the letters are the core of any language [2]. Moreover, the system presented here can be considered as a starting point for developing more complex systems.

There are two types of sign language recognition methods namely sensor-based and image-based. The first method is dependent on localized sensors or wearing specific gloves. The main benefit of this method is its ability to provide accurate information about signs or gestures such as the movement, rotation, orientation as well as positioning of the hands [6–10]. The second method uses different types of cameras. It is based on image processing which does not require equipment such as sensors. This approach only relies on the application of various image processing techniques and pattern recognition [5,10-14][19].

Sign language differs among countries. In addition, various sign

* Corresponding author.
E-mail addresses: 185105405@ybu.edu.tr (A. KASAPBAŞI), 195105116@ybu.edu.tr (A.E.A. ELBUSHRA), 185105406@ybu.edu.tr (O. AL-HARDANEE), a.yilmaz@maastrichtuniversity.nl (A. YILMAZ).

https://doi.org/10.1016/j.cmpbup.2021.100048
Received 21 June 2021; Received in revised form 13 October 2021; Accepted 29 December 2021
Available online 10 January 2022

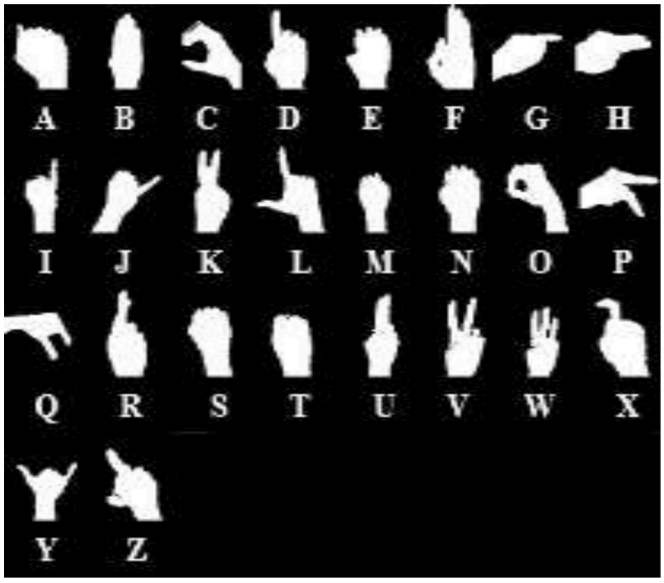

Fig. 1. Threshold hand poses of ASLA.

languages generally contain non-manual signs such as body gestures and facial expressions. They often require two hands or sequential movements for performing these signs. These issues increase the complexity in design of sign language recognition systems. To overcome this, researchers showed a specific interest in the sign language recognition systems [15].

In recent years, researchers have employed deep learning for sign language recognition systems. This involves the use of various methodologies and datasets which aim to improve the accuracy of the system. Various datasets are created because of many factors such as regional differences, type of images (RGB or Depth) and so on. Sign language differs from one region to another just like spoken languages and includes American sign language, Indian sign language, Arabic sign language, etc. [3,4,12,13,14,16,17,18]. Moreover, the type of images used for recognition systems depends on the camera that generates RGB [3,4,12,14,16,17,18] or depth images [4,13,18]. Furthermore, the methodologies used by different researchers that underlie the core of gesture recognition systems, vary from one system to another. Each study works to create and improve a different system to enhance its accuracy. Currently, there is no system which can deal with all conditions with high accuracy. Researchers have previously focused on CNNs with different parameters for sign language recognition systems due to their high performance in image classification [3,4,12,14,16,17,18]. Also, some studies use CNNs along with other methods to obtain more accurate results [16]. However, others have used methods such as SVM and PCANET [12,13,17]. Comparisons of CNNs with other methods have proved the superiority and ability of CNNs [12,17].

In this study, we have created an American Sign Language Alphabet (ASLA) dataset and developed a deep learning-based method for its recognition. The new dataset is designed to overcome a common challenge faced by researchers by improving the SLR system. Some challenges in the recognition of the letters using the ASLA dataset[1] exist due to variation in lighting and distance. Moreover, our dataset would improve the use of methods in the fields of machine learning and deep learning by the application of different methods using this dataset and comparison of the results. The images in the created datasets were captured with laptop and smartphone cameras. The proposed method primarily focuses on static hand gestures. In this study, a CNN-based deep learning method is proposed to create a robust and real-time ASL recognition system, because CNN as a deep learning technique has shown outstanding performance in tasks of image classification and pattern recognition [1,3,4]. The CNN model described here is evaluated with our ASLA dataset as well as datasets from other research [23,24].

The rest of the paper is organized as follows. The next section contains the methodology of this study. The methodology section provides detailed information about our dataset and other datasets used in this study as well as the architecture of the proposed deep learning system. The results section presents the findings of our study. Then, the next section discusses the results. Finally, the conclusions and future work is presented.

Methodology

A. Dataset

In this study, a custom dataset was constructed to interpret all the hand poses of the American sign language alphabet from gestures to labels. The dataset contains images and corresponding letters of ASLA. Fig. 1 shows the hand poses of the American sign language alphabet, which consists of 26 letters. In American sign language (ASL), each letter is depicted by a static sign made by using the hand to present it (excluding J and Z). The creation of the dataset was dependent on many factors such as illumination and the distance between the camera and hand which we adjusted to improve the performance of the convolutional neural network model. While in other datasets, the distance of the hand from the camera was reported to be fixed such as 0.5 m, 0.75 m or 1 m. Our dataset contains images varying 0.5 m, 0.75 m and 1 m hand distance. Furthermore, our dataset was collected at different times of the day specifically to include different illumination conditions. This feature is not available in the other ASL alphabet datasets. Therefore, our dataset overcomes this specific challenge for other researchers. Moreover, both the Z and J letters are created by a small movement that represents the letter, but in our study, we represent these letters by a static gesture at the start or end of the sign. This allowed us to include all the alphabet in the current dataset and obtain a complete set. Also, the required movement for these letters (Z and J) requires a separate neural network structure for achieving the recognition operation. However, our method is effective in recognizing the J and Z letters.To the best of our knowledge, there is currently no study that has reported such a method.

In the field of deep learning, when a new dataset is created, it may be considered a new contribution to the field mainly because each dataset has its specific features to improve existing models. However, the availability of several datasets often creates more challenges that require solutions. Therefore, the creation of a custom dataset with special conditions may be considered as a new contribution in the field of sign language interpretation.

In addition to the custom-constructed dataset, this study evaluates two more datasets using the proposed convolutional neural network model. In each evaluation, the corresponding dataset is separated into training, validation, and testing sets. The dimensions of the images are 64×64. The first dataset includes 52,000 images of the American sign language alphabet [23]. Also, the second dataset consists of 62,400 images [24]. Finally, the custom ASLA constructed in this study consists of 104,000 images. All datasets depend on the principle of image thresholding to generate binary images. Also, the datasets are split according to the following ratios 70:15:15; 70% for training sets, 15% for validation sets and 15% for testing sets.

B. System architecture

Convolutional neural network

A convolutional neural network (CNN) is one of the most commonly used deep learning methods to analyze visual imagery. CNN involves less preprocessing compared to other image classification algorithms. The network learns the filters that are normally hand-engineered in other systems. The use of a CNN reduces the images into a format that is easier to process while preserving features that are essential for making accurate predictions. There are four types of operations in a CNN:

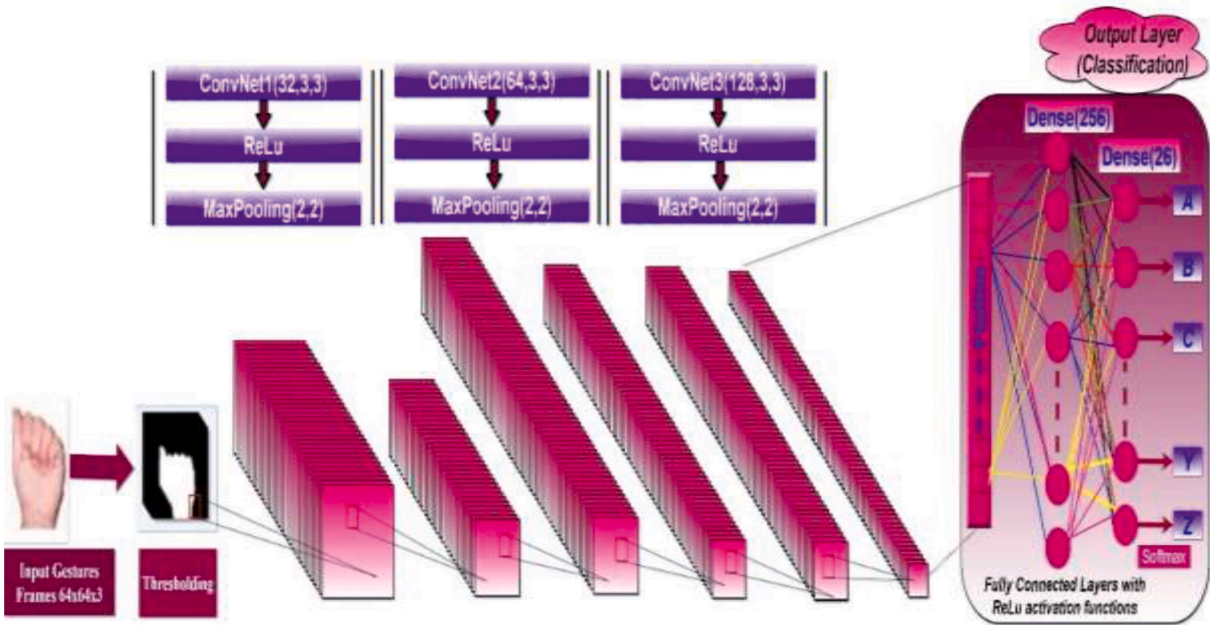

Fig. 2. The Proposed CNN model.

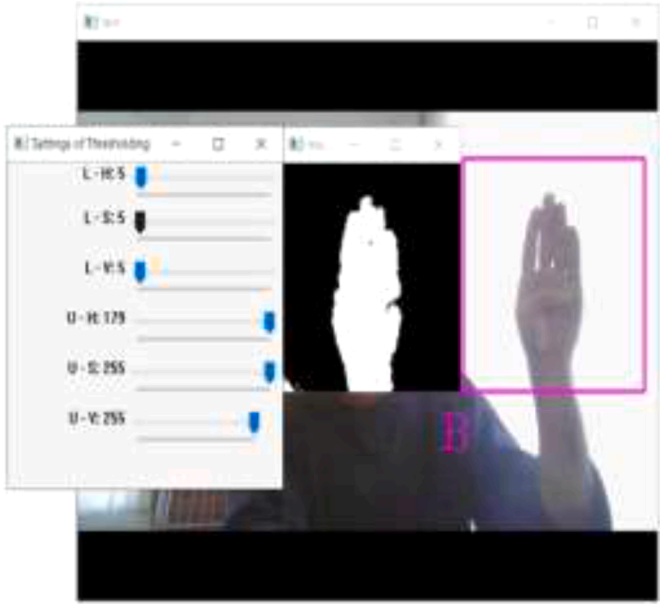

Fig. 3. The proposed prediction interfaces.

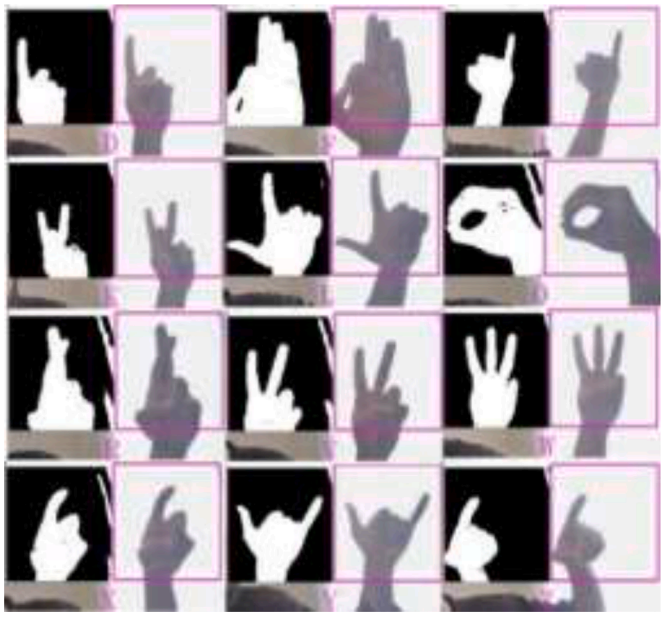

Fig. 4. Some samples of ASLA letters of this search.

convolution, pooling, flattening, and fully connected layers [20].

The convolution layer usually captures low-level features such as color, edges, and gradient orientation. The pooling layer decreases the spatial dimension of the convolved feature. This operation reduces the required computational time for dealing with the data through dimensionality alleviation. Furthermore, it has the advantage of maintaining dominant features that are positionally and rotationally invariant during the model training process. After the input image has been processed the higher-level features may be used for classification. Therefore, the image is flattened into a 1-D vector. In CNN, the flattened output is supplied to a fully connected layer. After training, using SoftMax classification, the model can provide probabilities of prediction of objects in the image [21].. Backpropagation is used to train the network. In this study, the system is implemented by using Keras, Tensorflow and OpenCV libraries.

The structure of the proposed CNN

The CNN model designed in our study consists of multiple layers. Fig. 2 illustrates the proposed structure of the CNN which consists of an input layer to input the images with 64 × 64 × 3 dimensions; this represents the size of the sign language frames that are taken as input into the system. The feature extraction part comprises three convolutional layers (Conv1, Conv2, Conv3). The convolution filter dimensions in each layer are 3 × 3. The number of filters is 32 filters for ConvNet1, 64 filters for ConvNet2 and 128 filters for CovNet3. Each convolution operation is followed by rectified linear units (ReLu). After ReLu, MaxPooling is applied with 2 × 2 dimensions. Pooling aims to prevent the loss of important information when the feature is represented. After the convolutional stage, flattening is applied for the classification stage. The classification stage is implemented with fully connected layers followed by a ReLu activation function and one SoftMax output layer [22].

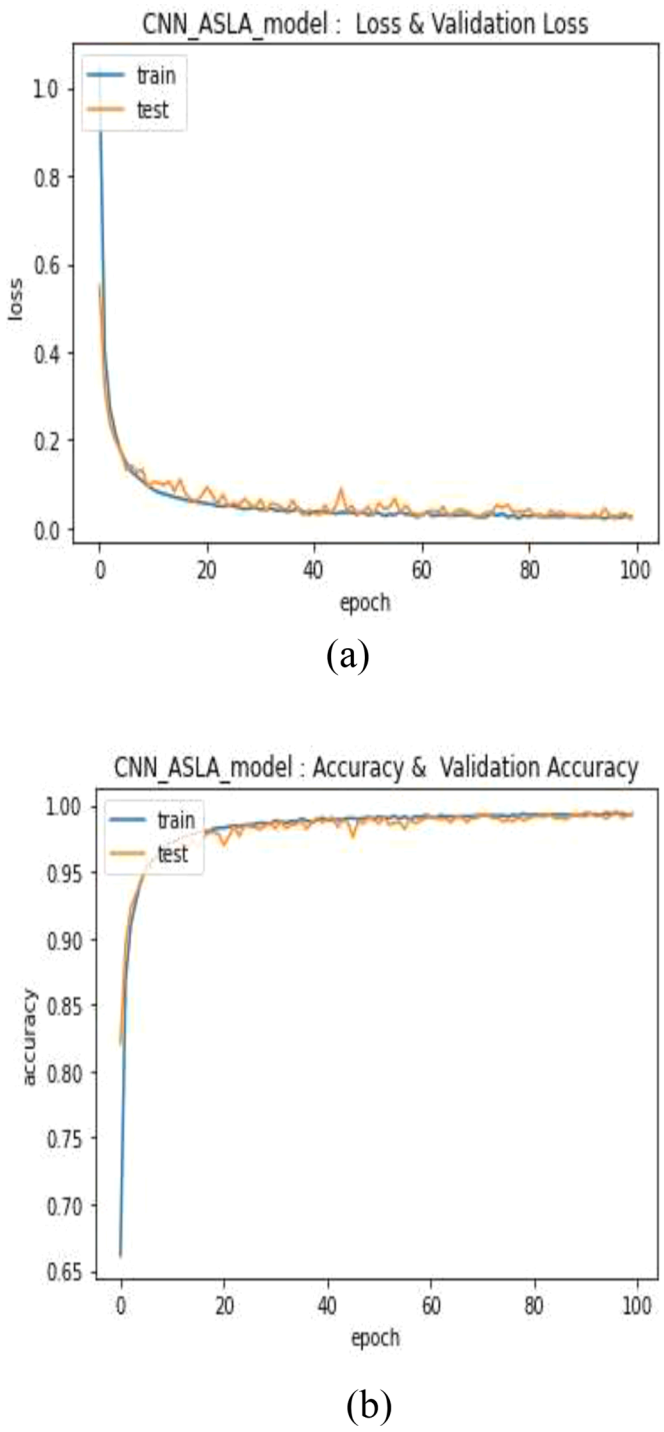

Fig. 5. (a) Training Loss and Validation Loss, (b) Training Accuracy and Validation Accuracy for the first dataset [23].

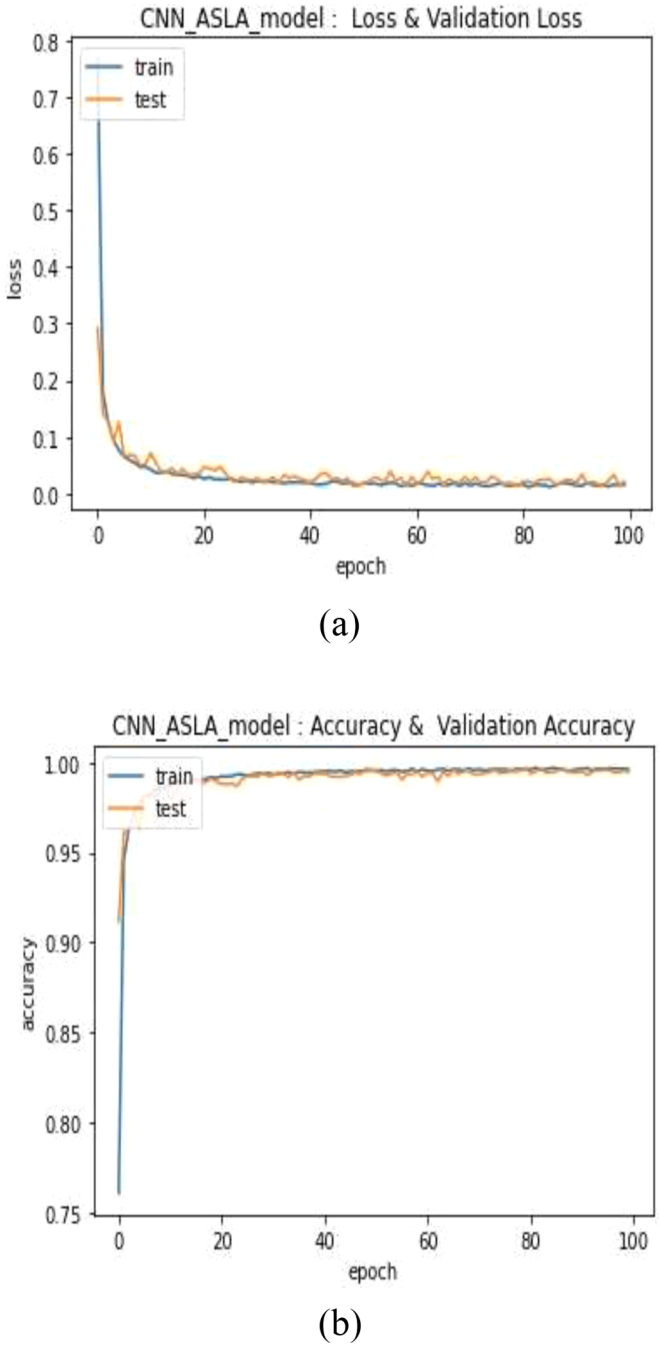

Fig. 6. (a) Training Loss and Validation Loss, (b) Training Accuracy and Validation Accuracy for the second dataset [24].

C. Application

Several researchers have tried to use deep learning for improving sign language recognition systems. G. Anantha Rao et al. (2018) proposes a CNN for the recognition of the gestures of the Indian sign language [3]. They used a continuous capture method from smartphone cameras in selfie mode and built a mobile application. They also generated a dataset from five different subjects performing 200 signs in 5 various viewing angles under different backgrounds due to the lack of availability of datasets on mobile selfie sign language. Moreover, they trained a CNN with 3 various sample sizes. Each one consisted of numerous subjects and viewing angles. The remaining 2 samples were utilized for testing the trained CNN. As a result, the researchers obtained 92.88% accuracy in comparison to other classifier models reported on the same dataset [3].

R. Daroya et al. (2018) reported a method which is CNN inspired called Densely Connected Convolutional Neural Networks (DenseNet) to classify RGB (Red_Green_Blue) images of static letter hand gestures in Sign Language [4]. They also utilized a web camera in real-time to achieve their work. DenseNet has advantages including the alleviation of a vanishing gradient. The advantages of DenseNet have been used widely for classification tasks. This proposed deep network is useful for sign language classification tasks and has an accuracy of 90.3%. In addition, they use both depth images and RGB images. V. Jain et al. (2021) created a system to recognize ASL by using both Support Vector

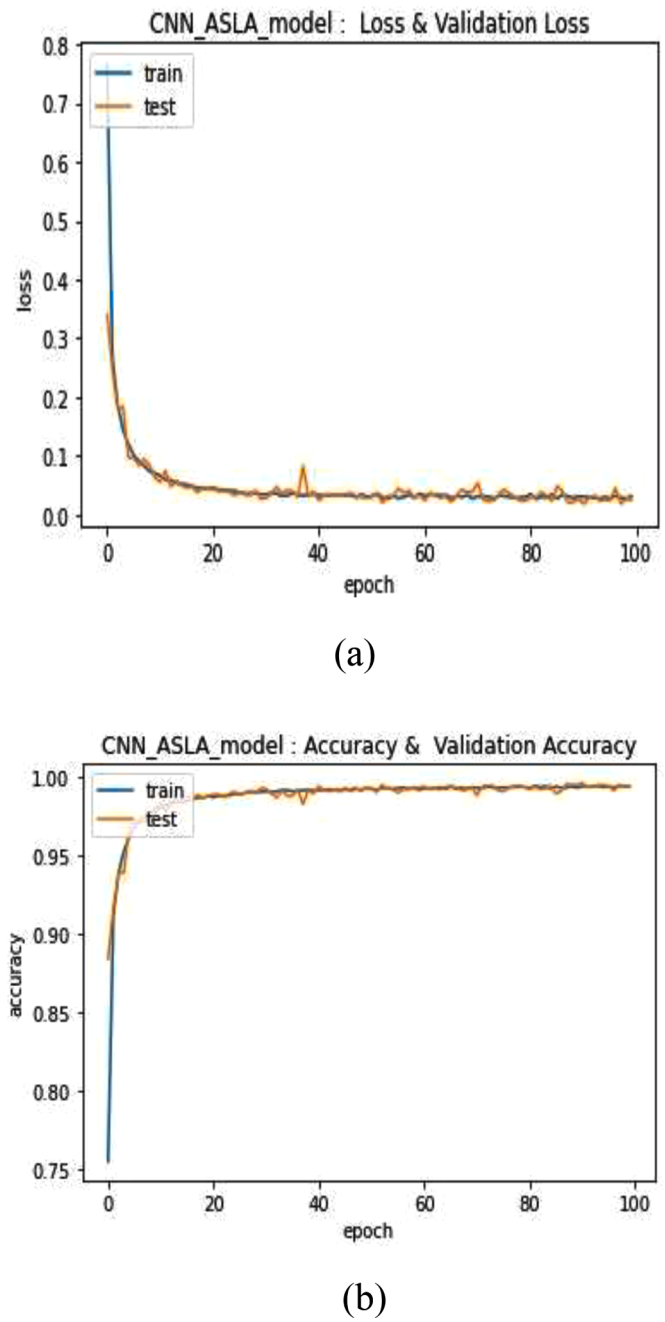

Fig. 7. (a) Training Loss and Validation Loss, (b) Training Accuracy and Validation Accuracy, for our dataset (ASLA).

Machine (SVM) and Convolutional Neural Network (CNN) [12]. They then compared the accuracy of the two methods. They found that the single layer CNN yielded 97.344% accuracy and 98.581% accuracy was reported for a two-layer CNN. In contrast, the SVM gave 81.49% accuracy when combined with the "Poly" kernel. They further aimed to enhance the accuracy of the CNN by determining the best size of the filter. S. Alyl. et al. (2017) designed a system for alphabetic Arabic sign language recognition by using intensity and depth images, which were obtained from a SOFTKINECTM sensor [13]. They utilized a PCANet method to learn local features from intensity and depth images and a linear support vector machine classifier to recognize the extracted features. The performance of the suggested system was assessed on a dataset of real images captured from multi-users. They implemented the tests in two ways; the first one utilized intensity and depth images separately, the second one utilized a combination of intensity and depth images. The acquired results revealed that the second test produced an accuracy of 99.5%. P. Kurhekar et al. (2019) present a system that can learn signs from videos by processing the video frames under a minimum disturbed background [14]. After that, the obtained sign is converted to readable text. This system utilizes a CNN which depends on a ResNet-34 CNN classifier, Fast.ai and OpenCV libraries for webcam inputs and displaying the predicted signs. They obtained a model with an accuracy of 78.5% on the presented testing set. K. Bantupalli and Y. Xie (2019) implemented a vision-based application that provides sign language translation into text in their work [16]. They aimed to aid people with hearing disabilities to communicate with hearing individuals. Their suggested model extracts temporal and spatial features from video sequences. Besides, they utilize a CNN for recognizing spatial features and a Recurrent Neural Network (RNN) to train on temporal features. They train the CNN and RNN models independently. They used the cross-entropy cost Adaptive Moment Estimation (Adam) function to minimize loss. Also, they double the size of their test dataset to gather more information from predictions about the model. In this system, the utilized dataset is the American Sign Language dataset, and the accuracy is 91% for 150 signs at the output of the softmax layer.

H. B.D Nguyen and H. Ngoc Do (2019) reported a sign language fingerspelling alphabet identification system [17]. In this study, three techniques are used for improving the system. In the first model, they utilize the combination of Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP) features and multi-kernel multi-class Support Vector Machines (SVM) to increase the special attributes of each feature type to the maximum point. Specifically, the researchers use 24 alphabetical symbols presented by static gestures, but they exclude two motion gestures which are the J and Z letters because of the dynamic structure of these letters. HOG and LBP features of each gesture are extracted from training images. After that, they train these extracted data by multi-class SVMs. The researchers also utilize an end-to-end CNN architecture for comparison. The combination of CNN as a feature descriptor and SVM generated an acceptable result. The feature-kernel pairs: HOG and RBF, LBP and Polynomial have an average recognition rate higher than two methods that utilize only one feature (98.36%). The CNN and CNN-SVM models provide 97.08% and 98.30% accuracy, respectively, which provides evidence that by executing CNN as a standalone feature extract, better results could be obtained than utilizing other CNN architecture. Despite the lower accuracy of the CNN-SVM model than that of the HOG-LBP- SVM model, it has a better opportunity when facing overfitting. S. Ameen and S. Vadera (2016) improve a convolutional network which they achieved by applying it to the problem of fingerspelling recognition for ASL [18]. This method improved the CNN model's classification performance in fingerspelling images by using both image intensity and depth data, introducing the applicability of deep learning for interpreting sign language and development. Their results showed that the improved model had 82% precision and 80% recall. Furthermore, they evaluated the confusion matrix and identified difficulties in classifying some signs.

In this study, we also developed an application for American sign language alphabet recognition in real-time. The OpenCV library was used to capture video due to its performance and robustness. The capture resolution is 640 × 480. Moreover, three windows appear when the application is running in prediction mode. Fig. 3 shows three windows of the application. The first window is the image capture window which contains a predefined region with dimensions of 196 × 196. The user is required to perform the signs in the predefined region. The application captures images with 64 × 64 dimensions which is necessary for the CNN model. After capturing the image from the predefined region, the thresholded image is displayed in the preview window. Then, the application saves the threshold display capture images to supply to the CNN model for classification. Finally, the predicted character appears under the predefined region of the image capture window. The control window with track bars is used for controlling the threshold. Fig. 4 illustrates some samples which were predicted by the program for

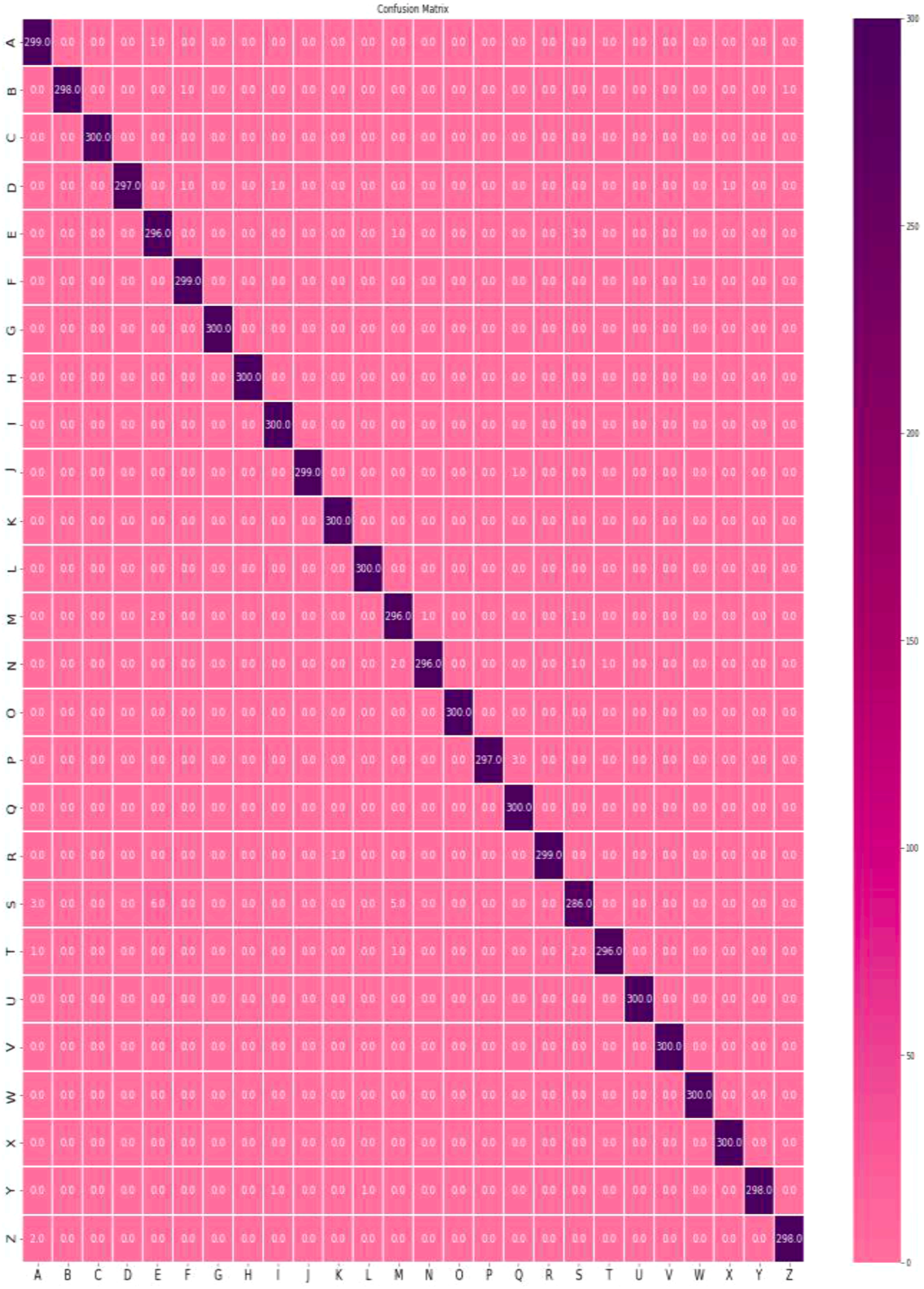

Fig. 8. Confusion matrix of test dataset for the first dataset in [23].

different American sign language Alphabet letters and the Z letter appears as a static letter at the start point of the required movement.

Results

Three separate American Sign Language Alphabet datasets were used to evaluate the proposed CNN model. These datasets are described in the methodology section. According to the results of the experiments, the proposed CNN model outperformed other neural network structures in certain metrics. The three datasets were applied to the proposed CNN for 100 epochs according to the following: Firstly, the training was executed for the first dataset [23] and the obtained accuracy was 99.41% with a 0.0204 loss. Secondly, the training was implemented to the second dataset [24], for which the obtained accuracy was 99.48% and the loss was 0.0210.

Finally, we trained the suggested CNN to our dataset (ASLA). Here,

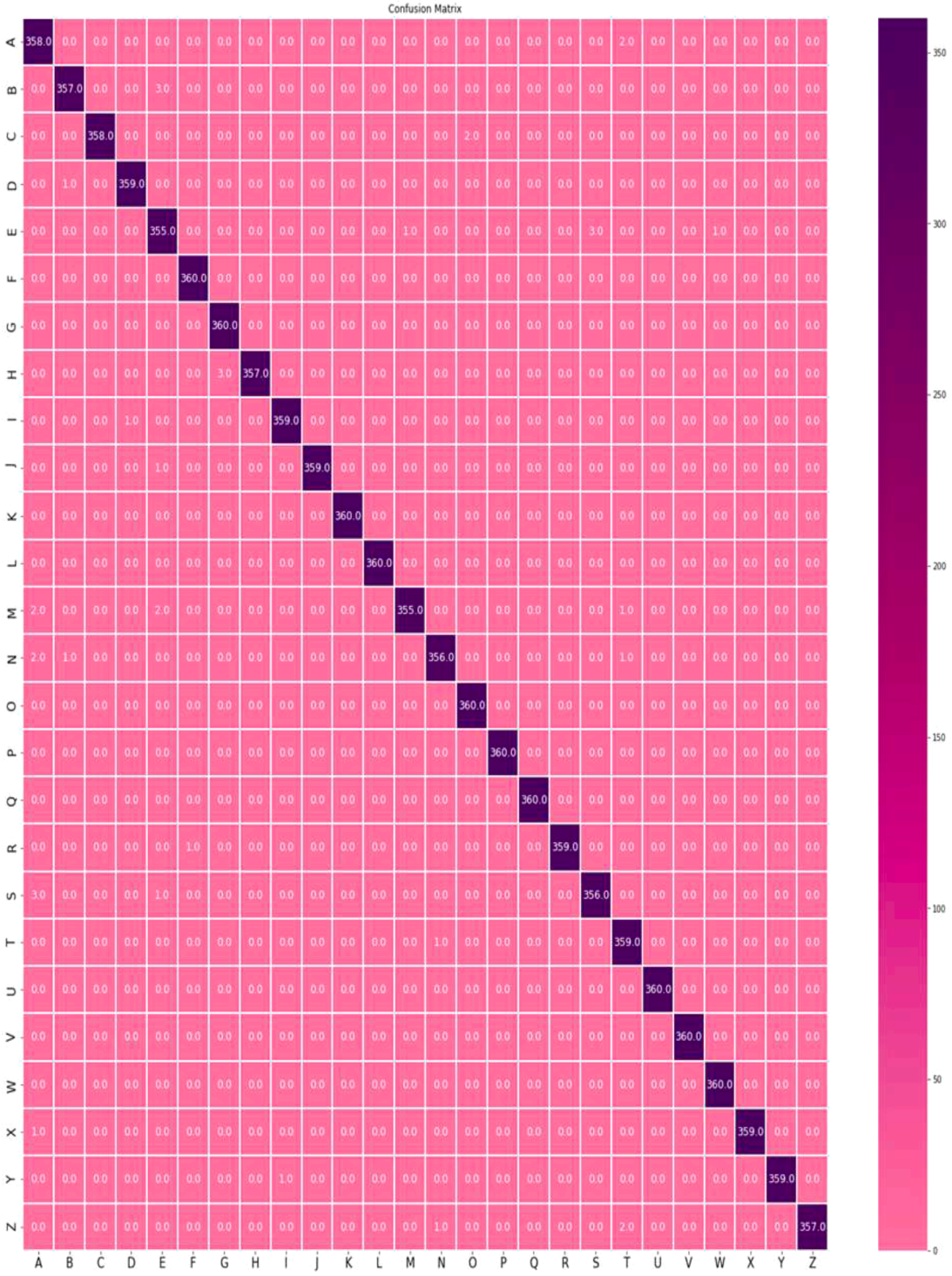

Fig. 9. Confusion matrix of test dataset for the second dataset in [24].

the obtained accuracy was 99.38% and the loss was 0.025. In this case, the dataset was larger than the two others. Figs. 5, 6, and 7 show the model losses which represent the training loss versus validation loss in part (a), and the model accuracy which shows the training accuracy versus the validation accuracy in part (b) for the first, second and our (ASLA) dataset sequentially. Moreover, Figs. 8, 9, and 10 represent the confusion matrix of the test dataset for the first dataset [23], second dataset [24] and ALSA dataset sequentially. The confusion matrices illustrate the high performance of the suggested model. Table 1 illustrates a comparison of the accuracies and losses the CNNs applied on the three different datasets.

Although the experimental results show that the first and second datasets have an accuracy higher than our dataset, the newly generated dataset is more challenging because it relies on many different conditions and factors. Furthermore, our dataset was the largest one among all the other datasets and contains 104,000 images which ultimately led to the superior prediction.

Discussion

This research presents a new dataset of the American sign language alphabet. The dataset which we have created takes into consideration

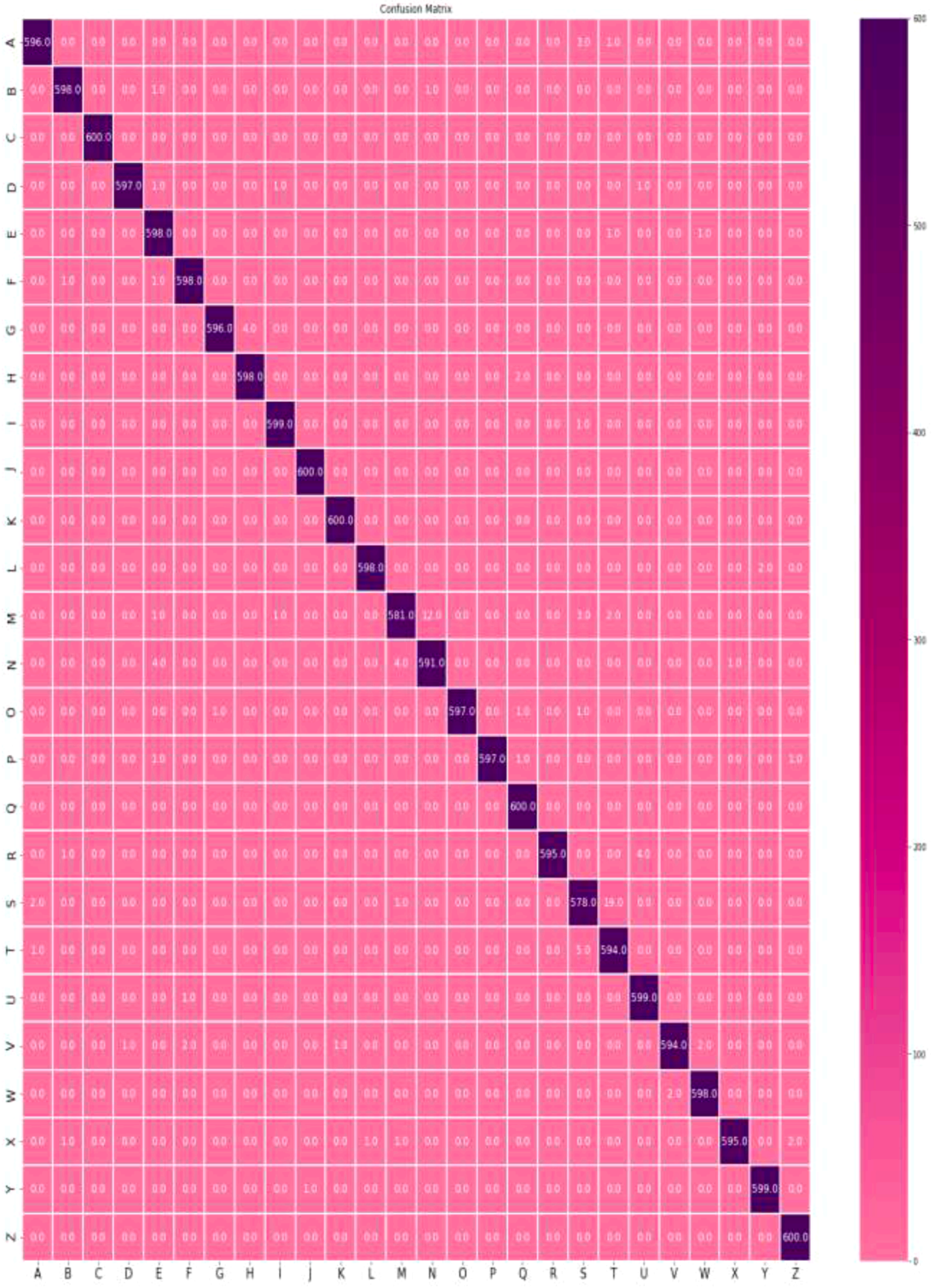

Fig. 10. Confusion matrix of test dataset for our dataset (ASLA).

Table 1
Comparison of the accuracies and for the various dataset.

The Proposed CNN Trained datasets	Accuracy	Loss
The first dataset [23]	99.41%	0.0204
The second dataset [24]	99.48%	0.0210
Dataset of this research (ASLA) [25]	99.38%	0.0250

various conditions such as lighting and distance, which makes it superior to other datasets which used non-variable conditions. This dataset is a new addition to other datasets in the field of SLR, which can assist researchers and students in developing SLR systems. Furthermore, our work presents a CNN that performs at a high accuracy under various conditions. Moreover, in our study, we have compared the accuracy of our dataset with those from other studies. Despite the different conditions and volume of the new dataset, it achieved 99.38% accuracy with

Table 2
Comparison of the accuracy between the suggested method and the methods in the other studies.

Method	Accuracy
CNN [3]	92.88%
DNN [4]	90.3%
CNN [12]	98.581%
SVM [12]	81.49%
PCANet [13]	99.5%
ResNet-34 CNN [14]	78.5%
CNN and RNN [16]	93%
CNN [17]	97.08%
CNN-SVM [17]	98.30%
Our CNN for three datasets	99.41% [23]
	99.48% [24]
	99.38% (ours)

better prediction and a small loss (0.0250). The comparison shown in Table 1 emphasizes the importance and contribution of our work and highlights the superior and competitive performance of our dataset. Table 2 compares the accuracy of our proposed approach with the accuracy of other methodologies used in other studies. Our method proves its ability to give superior accuracies for different datasets. For example, the methods used by several other researchers [3,4,12,13,14,16,17] reported accuracies that are seen in Table 2.

Conclusion and future work

Sign language is frequently used for communication by individuals who have a hearing impairment. As a result, the development of sign language recognition systems has a significant impact on these individuals. For this reason, a large amount of research has been attempted to develop a device that can automatically recognize the physical gestures of sign language. There have been numerous systems designed to interpret signs from images, but many problems are still faced by researchers in the field. Such challenges include the variation in size, position, shape and background of the hand, lighting, and the distance of the hand from the camera. Many studies have focused on the creation of sign language recognition systems by combining feature extraction methods with classification methods to identify hand poses. In this study, we developed a CNN architecture to interpret the American sign language alphabet. All layers have the same filtering window sizes of 3×3, which improves the speed and accuracy of recognition, but have an increased computational time. Additionally, max-pooling was used after each convolution layer. This study utilized three convolutional layers since increasing or decreasing the number of layers affects the performance of the neural network.

The CNN model that we have designed has three convolutional layers since we found this number of layers provided the best accuracy in empirical trials. Three datasets have been used for evaluation. Two datasets were from other studies to measure the loss and accuracy of the designed model for each dataset. The third dataset was created in-house in this study. This new dataset may support future research in the field of machine learning and deep learning to develop sign language recognition systems. The accuracy of the two previously acquired datasets is 99.41% with 0.0204 loss for the first dataset and 99.48% with 0.0210 loss for the second dataset, while accuracy of 99.38% with 0.025 loss was obtained with our dataset. In real world tests, these accuracies seem competitive, but the most accurate prediction was obtained by our dataset because our dataset is larger than others. Therefore, the CNN architecture that we have presented yields higher performance than previous SLR models.

This study can be improved by adding more images for more letters and words into the dataset. Also, more images can be added to improve accuracy and reduce loss. By the addition of new words and terms, the proposed system may be improved to predict a complete word. Additionally, these predicted words can be turned into speech by utilizing a text-to-speech engine. Other future work may be carried out aiming to convert the ASL to commands which robots or machines can understand and execute. In this situation, any person may stand in front of the robot and give commands to it by ASL.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] M. Mohandes, J. Liu, M. Deriche, A survey of image-based Arabic sign language recognition, in: 2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14), 2014, pp. 1–4, https://doi.org/10.1109/SSD.2014.6808906.
[2] C. Padden, D.C. Gunsauls, How the alphabet came to be used in a sign language, JSTOR 4 (2003) 10–33, https://doi.org/10.1353/sls.2003.0026.
[3] G.A. Rao, K. Syamala, P.V.V. Kishore, A.S.C.S. Sastry, Deep convolutional neural networks for sign language recognition, in *2018 Conference on Signal Processing And Commun. Eng. Syst. SPACES* (2018), 2018, vol. 2018-Janua, pp. 194–197, 10.1109/SPACES.2018.8316344.
[4] R. Daroya, D. Peralta, P. Naval, Alphabet sign language image classification using deep learning, in IEEE Region 10 Annual Int. Conference, Proceedings/TENCON (2019) vol. 2018-Octob, no. October, pp. 646–650, 10.1109/TENCON.2018.8650241.
[5] A. Thongtawee, O. Pinsanoh, Y. Kitjaidure, A novel feature extraction for American sign language recognition using webcam, in The 2018 Biomed. Eng. Int. Conference (2018) 5–9, https://doi.org/10.1109/BMEiCON.2018.8609933.
[6] J. Han, L. Shao, S. Member, D. Xu, J. Shotton, Enhanced computer vision with microsoft kinect sensor : a review, IEEE Trans. Cybern. 43 (5) (2013) 1318–1334, https://doi.org/10.1109/TCYB.2013.2265378.
[7] M. Mohandes, S. Aliyu, M. Deriche, Arabic sign language recognition using the leap motion controller, in IEEE 23rd Int. Symposium on Ind. Electronics (ISIE) (2014) 960–965, https://doi.org/10.1109/ISIE.2014.6864742.
[8] M.S. Kushwah, M. Sharma, K. Jain, Sign language interpretation using pseudo glove, in Proceeding of Int. Conference on Intelligent Commun. Control and Devices, Adva. Intelligent Syst. Comput. (2017) 9–19, https://doi.org/10.1007/978-981-10-1708-7.
[9] N. Pugeault, R. Bowden, Spelling it out : real – time ASL fingerspelling recognition university of surrey, in 3rd IAPR Asian Conference on Pattern Recognition (2011) 1114–1119, https://doi.org/10.1109/ICCVW.2011.6130290.
[10] V.N.T. Truong, A translator for American sign language to text and speech, in 2016 IEEE 5th Global Conference on Consumer Electronics (2016) 8–9, https://doi.org/10.1109/GCCE.2016.7800427.
[11] B. Kang, Real-time sign language fingerspelling recognition using convolutional neural networks from depth map, in 3rd IAPR Asian Conference on Pattern Recognition (2015) 136–140, https://doi.org/10.1109/ACPR.2015.7486481.
[12] V. Jain, A. Jain, A. Chauhan, S.S. Kotla, A. Gautam, American sign language recognition using support vector machine and convolutional neural network, Int. J. Inf. Technol. 13 (2021) 1193–1200, https://doi.org/10.1007/s41870-021-00617-x.
[13] S. Aly, B. Osman, W. Aly, M. Saber, Arabic sign language fingerspelling recognition from depth and intensity images, in 12th Int. Comp. Eng. Conference, ICENCO 2016: Boundless Smart Societies (2017) 99–104, https://doi.org/10.1109/ICENCO.2016.7856452.
[14] P. Kurhekar, J. Phadtare, S. Sinha, K.P. Shirsat, Real time sign language estimation system, in Proceedings of the Int. Conference on Trends in Electronics and Inf. ICOEI 2019 (2019) 654–658, https://doi.org/10.1109/ICOEI.2019.8862701, no. Icoei.
[15] M. Ahmed, M. Idrees, Z. Abideen, R. Mumtaz, S. Khalique, Deaf talk using 3D animated sign language, in 2016 SAI Comput. Conference (SAI) (2016) 330–335, https://doi.org/10.1109/SAI.2016.7556002.
[16] K. Bantupalli, Y. Xie, American Sign Language Recognition using Deep Learning and Computer Vision, in Proceedings - 2018 IEEE Int. Conference on Big Data, Big Data (2018) 4896–4899, https://doi.org/10.1109/BigData.2018.8622141, 2019.
[17] H.B.D. Nguyen, H.N. Do, Deep learning for American sign language fingerspelling recognition system, 2019 26th Int. Conf. Telecommun. ICT 2019 (2019) 314–318, https://doi.org/10.1109/ICT.2019.8798856.
[18] S. Ameen, S. Vadera, A convolutional neural network to classify American sign language fingerspelling from depth and colour images, Expert Syst. 34 (3) (2017), https://doi.org/10.1111/exsy.12197.
[19] K. Otiniano-Rodrıguez, E. Cayllahua-Cahuina, A. Araujo de A, G. Camara-Chavez, Finger spelling recognition using kernel descriptors and depth images, in IEEE 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (2015) 72–79, https://doi.org/10.1109/SIBGRAPI.2015.50.
[20] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, N U R E 521 (May 2015) 436–444, https://doi.org/10.1038/nature14539.
[21] K. He, J. Sun, Deep residual learning for image recognition, in IEEE Conference on Comp. Vision and Pattern Recognition (2016) 770–778, https://doi.org/10.1109/CVPR.2016.90.

[22] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in IEEE computer Vision and Pattern Recognition (2017) 2261–2269, https://doi.org/10.1109/CVPR.2017.243.
[23] R. Poudel, (2018, Jul 2), GitHub, "Simple-sign-language-detector". Available: https://github.com/rrupeshh/Simple-Sign-Language-Detector, (accessed 9 Feb 2021).
[24] D. Saha, (2018, May 9). GitHub, Sign-Language (Version1) Available: https://github.com/evilport2/sign-language, (accessed 9 Feb 2021).
[25] A. KASAPBAŞI, O. AL-HARDANEE, A.E. AHMED ELBUSHRA, A. YILMAZ, (2020, January 10), GitHub, "Recognition American sign language alphabets by CNNs". Available at: https://github.com/Ahmed-KASAPBASI/Success_Team_ASL.