2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT) | 978-0-7381-2323-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICAICT51780.2020.9333456

Classification of Sign Language Characters by Applying a Deep Convolutional Neural Network

Md. Mehedi Hasan*, Azmain Yakin Srizon†, Abu Sayeed‡ and Md. Al Mehedi Hasan§
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh
Email: *mmehedihasann@gmail.com, †azmainsrizon@gmail.com, ‡abusayeed.cse@gmail.com, §mehedi_ru@yahoo.com

***Abstract*—Having a massive community of almost 466 million deaf-mute people all over the world, sign language recognition has always fascinated researchers to develop sophisticated models that can successfully recognize sign languages. Because of not being a universal language, sign language differs in terms of languages and communities. Previously, various researches have been conducted on different sign languages. In this study, we considered the Sign Language MINST dataset. Previously, different classifiers like support vector machine, random forest, multilayer perceptron, etc. have been introduced for sign language recognition. Recently, shallow CNN and Capsule Networks have obtained better results. Therefore, in this research, we proposed a deep convolutional neural network model to achieve the successful identification of the sign linguistics alphabets. After implementing the model, we produced an overall accuracy of 97.62% and comparison with previous researches revealed that our proposed model outperformed all previously introduced models.**

***Index Terms*—American Sign Language, Alphabets Recognition, Deep Convolutional Neural Network**

I. Introduction

An inadequate or complete inability to hear is acknowledged as hearing impairment or hearing loss which may happen in the individual ear or both ears [1], [2], [3]. Hearing impairment may be prompted by several circumstances including genetics, aging, vulnerability to noise, several infections, birth complexities, injury of ear, and some medicines or venoms [2]. Till 2013, hearing impairment affected approximately 1,100 million personalities to some extent [4]. It has prompted an inability in almost 538 million persons and led to critical disabilities in approximately 124 million persons [2], [5], [6]. To overcome the connection gap with deaf-mute individuals, sign language is utilized which is the most practiced literature in the deaf-mute community to interact with people and yield opinions [7], [8]. Deaf-mute is a phrase that is practiced historically to recognize an individual who is either deaf or both deaf and cannot speak as well, and in both circumstances, sign language is the method of communication for them. Sign language is a language that utilizes visual-manual approaches to communicate or convey meaning. Like natural languages, sign languages possess their grammar and vocabulary [9]. Despite having significant similarities between sign languages, they are not universally the same and mutually signed [9].

Being stated that, previously, many kinds of research have been conveyed to successfully recognize sign languages of various communities all over the world [10], [11]. For example, a study of Indian sign language recognition gained 93% overall recognition accuracy by using enhanced skin and wrist discovery algorithms [12]. Researches on Spanish sign language proposed an overall accuracy of 96% [13]. However, in this research, we specifically focused on American sign language character recognition. Having 250,000 to 500,000 persons in the deaf community of Americans and some Canadians who utilize American sign language, this was an obvious opportunity for research [14]. Previously, various researches have contributed significant discoveries in the area of American sign alphabets identification. One of the studies suggested the identification of alphabets by using the Support Vector Machine classifier by utilizing histogram of gradients (HOG) features extracted from real-time hand gesture images [15]. However, the problem arises with the increase of dataset as it also increases training time. Another research discussed the variation in performance for hand gesture recognition by applying different methods, for instance, Random Tree, Naïve Bayes, C4.5 (J48), NNge, Random Forest, ANN (Multiplayer Perceptron), and SVM (Linear and RBF Kernel) [16]. However, a study suggested a deep convolutional neural network for recognizing American sign characters and outperformed previously discovered high-performance approaches like HOG + SVM (Liner Kernel), Random Forest, SVM (Linear Kernel), and Multilayer Perceptron (2 Hidden Layers) [17]. Recently, another study suggested 95.08% accuracy by applying Capsule Networks and introduced the comparison among performances of LeNet, CapsNet, and CapsNet with augmentation [18].

In this research, we inaugurated with the Sign Language MNIST dataset [18]. To classify the characters more precisely, we developed and introduced a novel deep convolutional neural network (CNN) model. With the support of our proposed model, the most significant features were obtained to generate a precise recognition of the characters of the American sign language dataset. The models have been executed on Keras framework that operated on top of Tensorflow. Our proposed model achieved an overall test accuracy of 97.62% that outperformed all previous researches.

II. Materials and Methods

A. Dataset Description

To evaluate our model on sign language recognition, we have adopted the Sign Language MINST dataset for American

978-0-7381-2323-3/20/$31.00 ©2020 IEEE

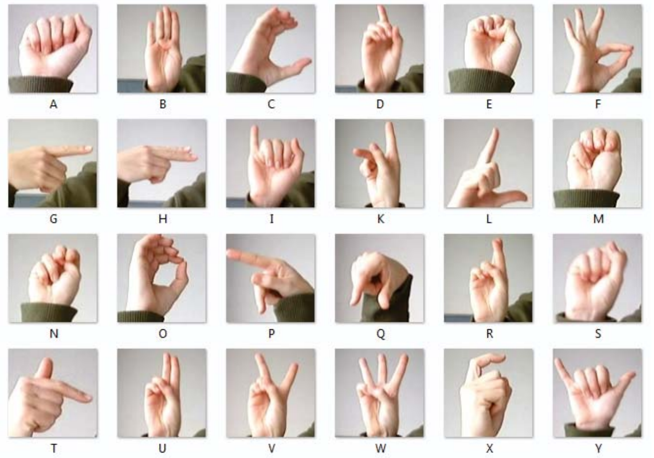

Fig. 1: This figure illustrates signs for American alphabets at a glance.

TABLE I: Comparison of overall accuracy among Proposed Model and previous researches

Classifier or Model Name	Overall Accuracy
Shallow CNN [17]	95.26%
HOG+SVM (Linear Kernel) [15]	90.71%
Random Forest [16]	65.57%
SVM (Linear Kernel) [16]	79.83%
MLP (2 Hidden Layers) [16]	75.68%
LeNet [18]	82.19%
CapsNet [18]	88.93%
CapsNet Augmented [18]	95.08%
Proposed Model	**97.62%**

sign language which is publicly accessible in Kaggle [18]. Among 26 alphabets, the American sign language dataset consists of 24 alphabetical gestures. As J and Z are motion-based alphabets and as we considered working with static images only, J and Z were excluded in the dataset. The dataset contains a total of 27,455 and 7,172 training and testing examples of 24 alphabetical gestures. From the training images, 30% of samples were chosen as validation samples to tune the model. Figure-1 showcases a glance of the American sign language dataset.

B. Proposed Deep CNN Architecture

Convolution Neural Network or CNN is one of the most traditional profound neural systems in the territory of picture recognition. CNN earned its notoriety after its remarkable execution in the ImageNet challenge [19]. A convolutional neural system has a few advantages over different kinds of neural systems in light of its innate creation or layers.

CNN has basically three kinds of layers: convolutional, pooling, and fully connected layers. In the convolutional layer, the information lattice gets duplicated with a few convolutional parts or channels to deliver an element map that clarifies what kind of highlight endures in a picture. In the pooling layer, interpretation and scaling fluctuation are given which likewise diminish the volume of highlight maps. At last, there exists the fully connected layer following all former convolutional and pooling layers which clarify what kind of attributes exist in the image and what kind of qualities don't exist in the image. These three kinds of layers are the fundamental layers of the Convolutional Neural Network. In addition, CNN has likewise fewer boundaries contrasted with other profound learning system structures.

In this study, a total of six convolutional layers and three pooling layers had been utilized. ReLU activation function was applied in the convolution layer. In the first two convolutional layers, 64 convolutional filters were used which was extended in the later-introduced two convolutional layers to 128 to get more deep features. Finally, in the latter two convolutional layers, 256 convolutional filters were employed in order to extract even more deep characteristics in the image. A 3x3 size filter was implemented in all the convolution layer and Max pooling was practiced for the pooling layer. Finally, the feature plucked from the picture was transported toward the fully connected layer. The ReLU activation function was utilized in the hidden layer too. There were 256 neurons in the hidden layer which recode mapping between the inputs from the fully connected layer and output from output layer. The output layer consisted of a total of 24 nodes with softmax [20] activation function as in the classification dilemma, we are dealing with 24 characters of American sign language. Figure-2 represents the design of the proposed CNN architecture.

III. Experimental Analysis

A. Preprocessing

First of all, all the RGB images were converted into grayscale images. Because of providing images to a convolutional neural network, a heavy preprocessing of the images was skipped as CNN is a powerful network that can detect useful features from raw images. However, two stages of preprocessing were performed on the American sign language dataset. As the raw images of the alphabets of American sign language were different in resolution, we had to resize the image into a fixed size. We resized the images into 28x28 as most of the images were in this range. Moreover, we encoded the class levels in order to align with the nature of the output of our proposed network.

B. Design of Experiment

The model was run for 50 epochs having a 256 batch size as after that the validation loss became nearly constant for the rest of epochs. 'Adam' optimizer including the learning rate of 0.0001 was practiced to maximize error function. A categorical cross-entropy function was applied for the error measurement. For bypassing overfitting, dropout technique was practiced.

C. Result Analysis

First of all, the dataset was split into 70% and 30% where 70% of the dataset was utilized as the train set and the additional 30% of the dataset was utilized for creating validation set. The test set was supplied separately and had

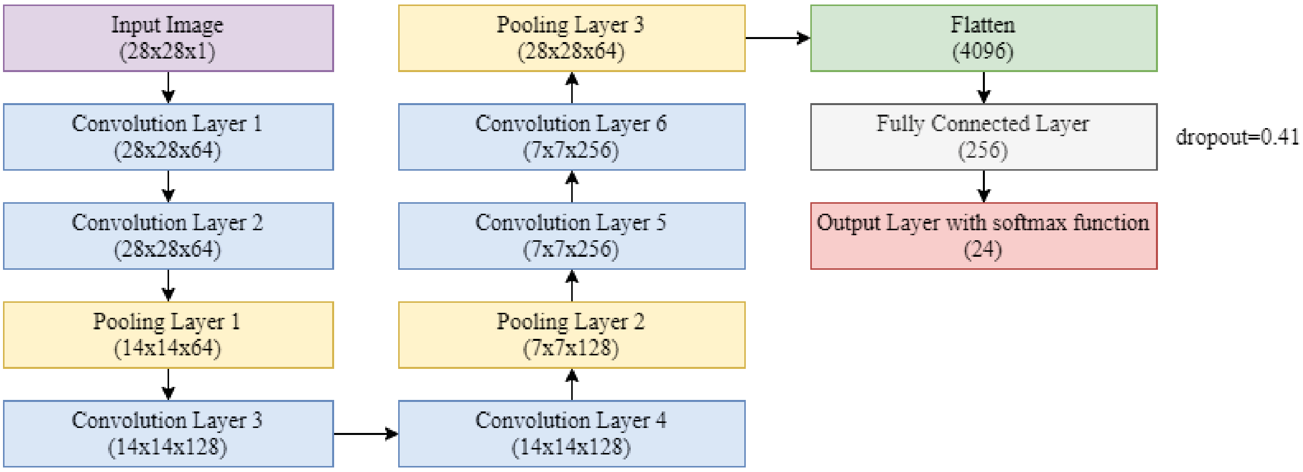

Fig. 2: Illustration of the architecture of our proposed CNN model.

TABLE II: Comparison of class-specific accuracy among Proposed Model and previous researches

Alphabet Ti	Proposed Model	Shallow CNN [17]	Linear SVM on HOG [15]	Random Forest [16]	Linear SVM [16]	MLP with Two Hidden Layers [16]
A	**1.0000**	**1.0000**	**1.0000**	0.9819	**1.0000**	**1.0000**
B	**0.9954**	0.9606	0.9884	0.8194	0.9005	0.8449
C	**1.0000**	0.8677	**1.0000**	0.9258	0.9839	0.9710
D	**1.0000**	0.9224	0.9184	0.8327	0.9592	0.8816
E	**0.9980**	0.9940	0.9820	0.8815	0.9659	0.9659
F	**1.0000**	**1.0000**	**1.0000**	0.8947	0.9150	0.9150
G	**0.9396**	**0.9397**	0.9310	0.7845	0.8994	0.8276
H	**0.9976**	**0.9977**	0.9495	0.8073	0.8303	0.8440
I	**1.0000**	**1.0000**	0.8646	0.4653	0.9028	0.6667
K	**0.9366**	0.9305	0.8731	0.4743	0.4773	0.5347
L	**1.0000**	**1.0000**	0.8134	0.8612	0.8421	**1.0000**
M	0.8934	**0.9467**	0.9086	0.5533	0.7487	0.6726
N	**1.0000**	0.9794	0.7491	0.3952	0.6942	0.5704
O	**0.9837**	0.9228	0.9146	0.6585	0.7398	0.8943
P	0.9914	**1.0000**	0.9625	0.9308	0.9395	**1.0000**
Q	**1.0000**	**1.0000**	0.8720	0.9573	0.9024	0.7439
R	0.7986	**0.9514**	0.8542	0.4375	0.7153	0.5694
S	**1.0000**	0.9106	0.9146	0.3862	0.6829	0.4634
T	**0.9153**	0.8347	0.7500	0.5968	0.6734	0.6331
U	0.9737	**0.9887**	0.8759	0.3383	0.6241	0.4586
V	**1.0000**	0.8353	0.9364	0.3237	0.7139	0.5000
W	**1.0000**	**1.0000**	0.8981	0.4126	0.8058	0.8058
X	**1.0000**	**1.0000**	0.9326	0.5281	0.6217	0.6966
Y	**0.9458**	0.8795	0.8825	0.4910	0.6205	0.7048

no influence over the training or the validation set. After applying the preprocessing steps described earlier, our proposed CNN architecture was implemented on the processed dataset. Figure-3 illustrates the train and validation efficiency of our proposed architecture. On the other hand, Figure-4 illustrates the loss of train and validation for our proposed design. For both of the figures, the blue line highlighted loss and accuracy for the train set, and the orange line indicated loss and accuracy of the validation set. Finally, this trained model was utilized to predict the overall accuracy of the test set provided separately with the American sign language dataset and our model obtained an overall accuracy of 97.62% which outperformed all other studies. Figure-5 represents the confusion matrix created from the actual classes and prophesied classes by our trained model. Table-1 showcases the difference in performance among our proposed model and models introduced in previous researches. For more convenience, in Table-2, we have shown the accuracy for separate classes and compared our results with the approaches found in previous researches which indicates higher accuracy for most of the

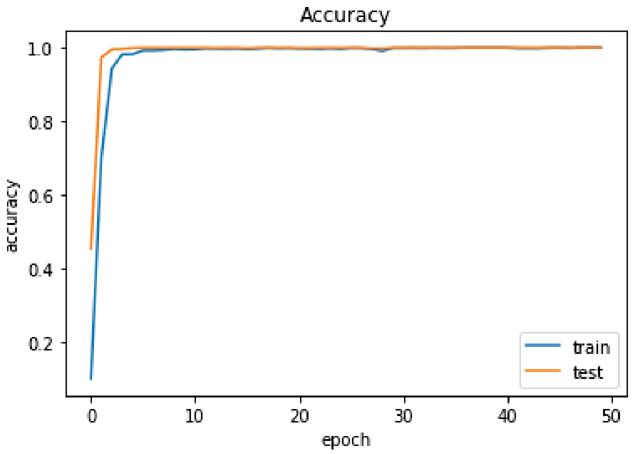

Fig. 3: Validation and training accuracy for the proposed CNN architecture while training period.

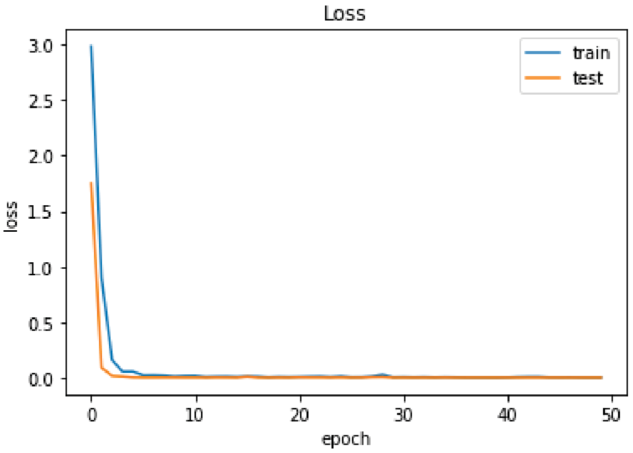

Fig. 4: Validation and training loss for the proposed CNN architecture while training period.

classes. From Table-1 and Table-2, we concluded that our model outperformed all previous researches by a significant margin in terms of American sign language detection.

IV. Conclusion

Sign language recognition has been an area of interest for researchers for a long time now. Previously, various researches on sign languages of different communities have been conducted for successful recognition of the sign language alphabets. In this research, we started with the Sign Language MNIST dataset of American sign language and proposed a deep convolutional neural network and calculated the overall test accuracy. A comparison with the previous researches revealed that our model outperformed all previously introduced models in terms of overall accuracy. We believe, in the future, many contributions can be possible in increasing the overall accuracy even more as well as creating a real-time American sign character recognition system.

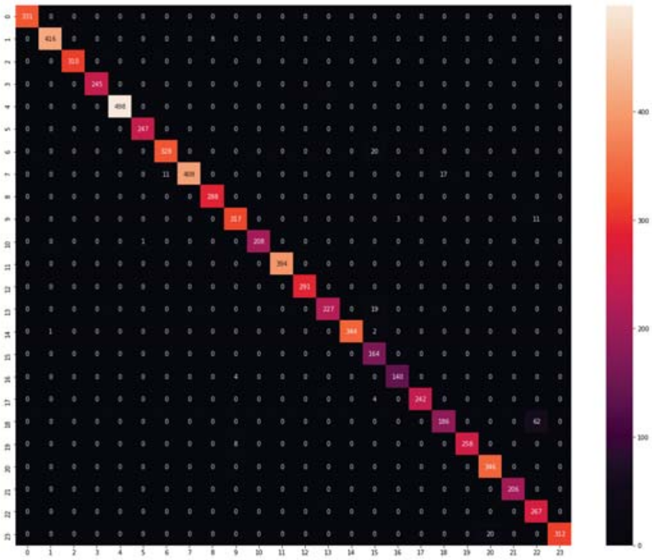

Fig. 5: This figure illustrates confusion matrix for the test set.

References

[1] E. Britannica and E. Britannica, “Inc., 2012,” *Encyclopædia Britannica Online. Website*, 2012.

[2] W. H. Organization *et al.*, “Deafness and hearing loss. fact sheet n 300. updated march 2015,” 2015.

[3] M. J. Cheok, Z. Omar, and M. H. Jaward, “A review of hand gesture and sign language recognition techniques,” *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 131–153, 2019.

[4] T. Vos, R. M. Barber, B. Bell, A. Bertozzi-Villa, S. Biryukov, I. Bolliger, F. Charlson, A. Davis, L. Degenhardt, D. Dicker *et al.*, “Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the global burden of disease study 2013,” *The Lancet*, vol. 386, no. 9995, pp. 743–800, 2015.

[5] W. H. Organization *et al.*, *The global burden of disease: 2004 update*. World Health Organization, 2008.

[6] B. O. Olusanya, K. J. Neumann, and J. E. Saunders, “The global burden of disabling hearing impairment: a call to action,” *Bulletin of the World Health Organization*, vol. 92, pp. 367–373, 2014.

[7] D. Bragg, O. Koller, M. Bellard, L. Berke, P. Boudreault, A. Braffort, N. Caselli, M. Huenerfauth, H. Kacorri, T. Verhoef *et al.*, “Sign language recognition, generation, and translation: An interdisciplinary perspective,” in *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 2019, pp. 16–31.

[8] J. Pu, W. Zhou, and H. Li, “Iterative alignment network for continuous sign language recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4165–4174.

[9] W. Sandler and D. Lillo-Martin, *Sign language and linguistic universals*. Cambridge University Press, 2006.

[10] R. Cui, H. Liu, and C. Zhang, “A deep neural framework for continuous sign language recognition by iterative training,” *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1880–1891, 2019.

[11] Y. Liao, P. Xiong, W. Min, W. Min, and J. Lu, “Dynamic sign language recognition based on video sequence with blstm-3d residual networks,” *IEEE Access*, vol. 7, pp. 38 044–38 054, 2019.

[12] K. Yadav, L. P. Saxena, B. Ahmed, and Y. K. Krishnan, “Hand gesture recognition using improved skin and wrist detection algorithms for indian sign,” *Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org*, vol. 9, no. 2, 2019.

[13] G. Saldaña González, J. Cerezo Sánchez, M. M. Bustillo Díaz, and A. Ata Pérez, “Recognition and classification of sign language for spanish,” *Computación y Sistemas*, vol. 22, no. 1, pp. 271–277, 2018.

[14] R. E. Mitchell, T. A. Young, B. Bachelda, and M. A. Karchmer, “How many people use asl in the united states? why estimates need updating,” *Sign Language Studies*, vol. 6, no. 3, pp. 306–335, 2006.

[15] E. Ohn-Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," *IEEE transactions on intelligent transportation systems*, vol. 15, no. 6, pp. 2368–2377, 2014.

[16] M. Lech, B. Kostek, and A. Czyżewski, "Examining classifiers applied to static hand gesture recognition in novel sound mixing system," in *Multimedia and Internet Systems: Theory and Practice*. Springer, 2013, pp. 77–86.

[17] D. Chakraborty, D. Garg, A. Ghosh, and J. H. Chan, "Trigger detection system for american sign language using deep convolutional neural networks," in *Proceedings of the 10th International Conference on Advances in Information Technology*, 2018, pp. 1–6.

[18] M. Bilgin and K. Mutludoğan, "American sign language character recognition with capsule networks," in *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. IEEE, 2019, pp. 1–6.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[20] R. A. Dunne and N. A. Campbell, "On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function," in *Proc. 8th Aust. Conf. on the Neural Networks, Melbourne*, vol. 181. Citeseer, 1997, p. 185.