2024 International Conference on Advancements in Power, Communication and Intelligent Systems (APCI) | 979-8-3503-6328-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/APCI61480.2024.10616504

Hand Gesture Classification for Individuals with Disabilities Using the DenseNet121 Model

Basel A. Dabwan
Department of Information System
AlBaydha University
AlBaydha, Yemen
baseldbwan@yahoo.com

Mukti E. Jadhav
Department of Computer Science
Shri Shivaji Science & Art College
Chikhali Dist. Buldhana., India
muktijadhav@gmail.com

Amol S. Gadkari
Department of Computer Science & IT
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, India
amolgadkari99@gmail.com

Yahya A. Ali
Department of Information System
Collage of Computer Science and Information Systems
Najran University, Kingdom of Saudi Arabia
yaali@nu.edu.sa

Soad M. almula
Department of Computer Science
College of Computer Science and Information Systems
Najran University, Kingdom of Saudi Arabia
smfadlmula@nu.edu.sa

Omar A. Ismil
Department of Information System
Preparatory in Najran University
Najran, Kingdom of Saudi Arabia
ismil8@gmail.com

Ashraf A. Mohammad
Department of Information System
Preparatory in Najran University
Najran, Kingdom of Saudi Arabia
aalzubir@nu.edu.sa

***Abstract*— The sign language functions as a method of communication for people who are deaf and mute, utilizing recognized signs or bodily gestures to convey meanings. It incorporates shapes, hand movements, directions, and facial expressions. A single sign not only represents a word but also communicates a particular tone. For many deaf individuals, verbal communication is not an option, and they may also face challenges in reading and writing. Consequently, the development of a sign language translation system, or more precisely, a sign language recognition (SLR) system, holds significant importance in their lives. SLR is highly sought after due to its potential to facilitate communication between those who are deaf and those who hearing individuals. This field represents a crucial area of research within the realm of human-computer interaction studies. To tackle this issue, we employed the DenseNet121 Model. Our approach involved using the DenseNet121 Model alongside a dataset representing the ASL Alphabet, encompassing 24 classes corresponding to English sign language letters (excluding the letters J and Z, which involve movement). The training dataset comprised 27,455 instances, while the test dataset consisted of 7,172 instances. During the training process, we allocated Using 80% of the dataset for training and setting aside the remaining 20% for testing. The outcomes of our proposed model were exceptionally encouraging, achieving an impressive accuracy rate of 97% and 96% validation accuracy. This success underscores the potential effectiveness of our model in automatically recognizing American Sign Language gestures, thereby enhancing facilitating communication and access for people who are hearing-impaired.**

Keywords— Deep Learning; Artificial Neural Network; Sign Language; Image processing; DenseNet121; Hand Gestures; individuals with disabilities.

I. INTRODUCTION

Sign languages provide a crucial methods of communication for people with hearing or speech impairments, offering an alternative way to interact. They also play a vital role as a universal communicative tool across different languages, especially beneficial in multicultural settings and emergencies. However, learning sign language can be challenging due to the numerous, sometimes similar, hand gestures required. This highlights the importance of developing an automated system for sign language recognition to improve accessibility.

This study addresses the need for a reliable detection model for the sign language alphabet using deep learning, which excels in image classification tasks. By utilizing the cameras commonly found in devices such as smartphones and laptops, our system processes image data alone, avoiding the complexity of depth information. This approach ensures easier integration and scalability, focusing on static hand gestures and excluding the dynamic gestures associated with

979-8-3503-6328-9/24/$31.00 ©2024 IEEE

the letters J and Z. Our model uses sign language of united stat of America as the basis for hand gestures.

The paper is organized as follows: The second section examines prior research and methodologies in this field. Section 3 outlines the suggested approach. Section 4 presents the results and discussions. Finally, Section 5 offers conclusions for the study.

II. RELATED WORK

Various approaches have been suggested to tackle the task of recognizing hand gestures in sign language. Early approaches, such as those mentioned in references [1] and [2], utilized Support Vector Machines (SVM) and parallel hidden Markov models, respectively. In another study [3], Sole et al. utilized Extreme Learning Machine (ELM) to classify static hand gestures that represent letters in the Auslan dictionary. However, while their initial findings were promising, the network's performance lacked generalization due to the limited test data collected on the same day.

In a different study [4], Kim et al. introduced a method leveraging They employed deep neural networks (DNNs) to classify frames in sequences of images depicting finger-spelled letters. Their approach achieved signer-independent recognition, capable of identifying hand poses across different users. They used Histogram of Gradients (HoG) image features [5] as input for the DNN. Similarly, in subsequent work [6], Kim et al. employed DNN in conjunction with HoG features, achieving optimal results through segmented Conditional Random Fields (CRF) with DNN.

Our research also aimed to achieve signer-independent image classification, yet unlike previous approaches, our classifier consists entirely of deep neural networks, encompassing feature extraction as well.

The study conducted by Pugeault et al. [7] introduced the utilization of depth pictures captured from a Microsoft Kinect device. They employed a multiclass random forest classification approach and assessed their method by varying inputs, including image-only, depth-only, and a combination of image with depth. Their optimal performance was achieved when combining depth with image inputs, and their system demonstrated sufficient speed for real-time recognation. Similarly, Kang et al. [8] utilized images with depth as inputs, exclusively without color images, employing a DNN classifier and achieving real-time processing rates. Our objective aligns with theirs in aiming for real-time processing, although we focus solely on color images due to the limited availability of depth sensors among users.

In recent years, CNN algorithm have made significant strides in the field of computer vision, particularly in tasks involving image recognition. Starting with AlexNet [9], which gained widespread recognition for winning the ImageNet Challenge: ILSVRC 2012 [10], subsequent architectures such as [11], [12], and [13] have continued to improve performance in various domains. This study aims to capitalize on advancements in deep learning techniques to create a robust and real-time classifier for finger-spelled sign language.

In arecent development, [14] introduced Dense Convolutional Network (DenseNet), a novel architecture designed to address challenges like vanishing gradients in deep networks. DenseNet achieves this by establishing dense connections between every layer in a feed-forward manner, promoting feature reuse and parameter efficiency. Given these advantages, our network model is based on the DenseNet architecture.

Moreover, in [15], researchers utilized pre-trained VGG16 and VGG19 models to efficiently translate sign language into written text. Trained on a comprehensive dataset comprising 27,455 training samples and 7,172 test samples, their approach achieved impressive accuracies of 97.5% with VGG16 and 96% with VGG19.

III. PROPOSED MODEL

The approach taken in constructing these models follows a practical and systematic methodology. Initially, we collected the required training data from the designated source, as elaborated later. Next, we underwent a preprocessing phase to refine the data for training purposes. Next, the data was divided into training and testing sets. Finally, we utilized the DenseNet121 model to classify sign language, as elaborated further below.

Fig. 1. Proposed Model.

A. Dataset

For this particular model, we utilized a dataset consisting of 27,455 images for training and 7,172 images for testing purposes. purposes. The dataset, obtained in CSV format, was sourced from Kaggle [16].

Fig. 2. Dataset for Sign Language.

B. Dataset Preprocessing

We performed several preprocessing steps to remove unnecessary elements, improve efficiency, and accelerate computations. These steps involved operations such as

resizing images, rescaling (by dividing each pixel value by 255), and partitioning splitting the dataset into training and testing subsets

C. *Division of training and testing data*

80% of the total dataset is allocated to the training dataset, while the remaining 20% is designated for the testing dataset.

D. *Extracting features and classifying with DenseNet121*

We developed sign language models by utilizing the DenseNet121 pre-trained model along with a dataset consisting of 24 classes representing sign language letters.

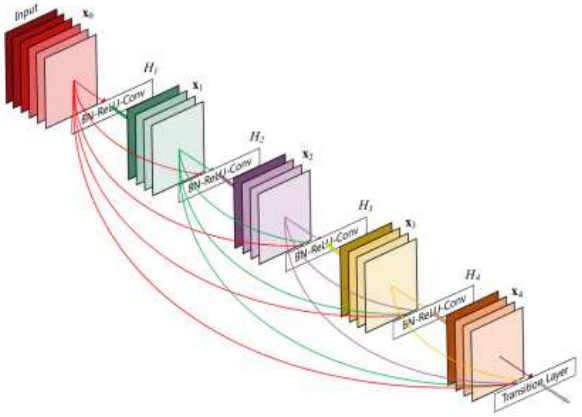

Fig. 3. The architecture of a Dense Block comprising 5 layers [14]

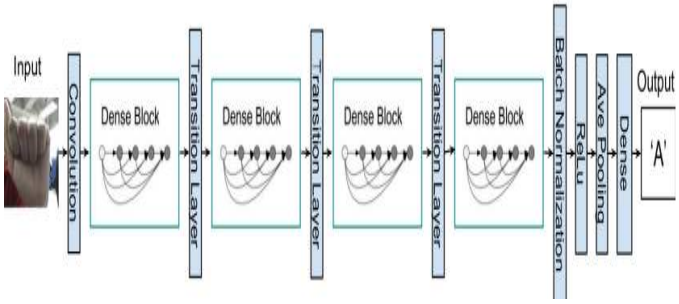

Fig. 4. The network structure employed for categorizing images into letters utilized four Dense Blocks [14]

IV. Outcome and Discussion

A. *OUTCOME*

Once the mentioned model was implemented, outstanding results were achieved, obtaining an accuracy of 97% in training and 96% in validation. as demonstrated below.

	precision	recall	f1-score	support
0	1.00	0.99	0.99	331
1	0.99	0.96	0.97	432
2	1.00	1.00	1.00	310
3	0.89	0.94	0.91	245
4	0.91	0.99	0.95	498
5	0.85	1.00	0.92	247
6	0.97	0.90	0.93	348
7	0.93	0.98	0.95	436
8	0.96	0.99	0.97	288
10	0.95	0.92	0.93	331
11	0.96	1.00	0.98	209
12	1.00	0.95	0.98	394
13	0.98	0.95	0.96	291
14	0.99	1.00	0.99	246
15	0.99	1.00	1.00	347
16	1.00	0.98	0.99	164
17	0.98	0.92	0.95	144
18	0.96	0.91	0.93	246
19	0.99	0.91	0.95	248
20	0.99	0.95	0.97	266
21	0.99	0.89	0.94	346
22	0.77	0.98	0.86	206
23	1.00	0.88	0.94	267
24	1.00	1.00	1.00	332
accuracy			0.96	7172

Fig 5. Our DenseNet121 Model's F1-score, recall, and precision

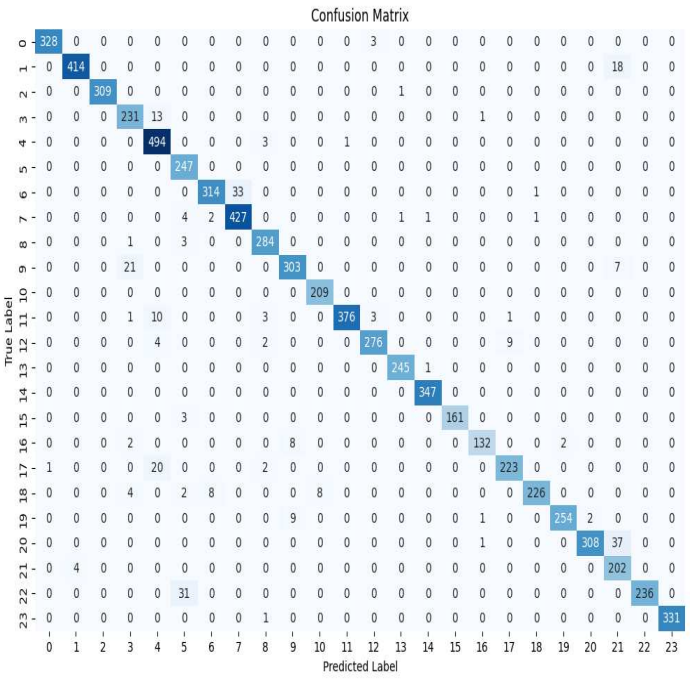

Fig. 6. Confusion Matrix of our DenseNet121 Model

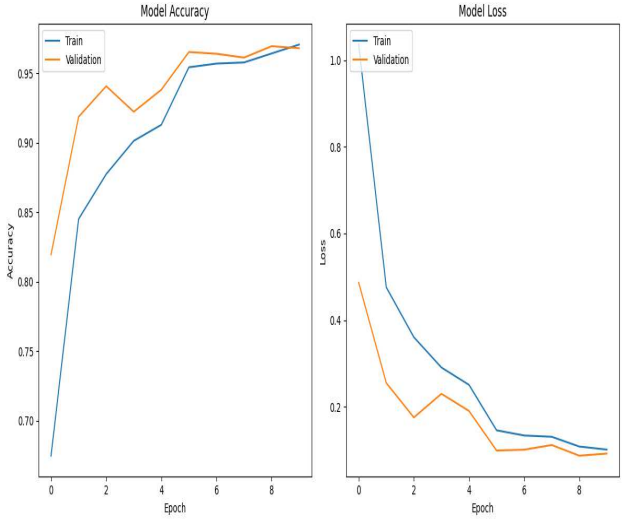

Fig. 7. Accourcy and Validation Accuracy of our DenseNet121 Model

B. *Comparison Outcomes*

After evaluating our developed models against existing pre-trained models for sign language detection, our findings demonstrate superior performance, as illustrated in the table provided.

TABLE 1. COMPARING THE PROPOSED MODEL TO EXISTING MODELS.

Sl. No	Model	Validation Accuracy (%)	Test Accuracy (%)
Model [17]	Inception-V3	88.45%	89.5%
Model [17]	ResNet-50	84.23%	85.3%
Model [17]	VGG-16	81.03%	82.0%
Model [18]	VGG-19	94%	96%
Model [19]	MobileNetV2	92%	91%
Model [20]	InceptionV3	96%	95%
Model [21]	ResNet50	89%	88%
Model [22]	VGG-16	95%	94%
Model [23]	DenseNet	90%	90%
Model [24]	EfficientnetB2	81%	80%
Proposed Model	DenseNet121	97%	96%

V. CONCLUSION

We created models for translating sign language, used by people who are deaf or mute into a format easily understandable by the general populace. Our goal in this endeavor is to assist people with disabilities in articulating their wants and requirements, thus promoting smoother communication with society. These models were built using datasets and trained with the DenseNet121 pre-trained model as previously mentioned. Significantly, one particular model yielded outstanding outcomes, achieving 97% accuracy in training and 96% accuracy in validation.

REFERENCES

[1] S. Naidoo, C. Omlin, and M. Glaser, “Vision-based static hand gesture recognition using support vector machines,” University of Western Cape, Bellville, 1998.

[2] C. Vogler and D. Metaxas, “Parallel hidden markov models for american sign language recognition,” in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 1. IEEE, 1999, pp. 116–122.

[3] M. M. Sole and M. Tsoeu, “Sign language recognition using the extreme learning machine,” in AFRICON, 2011. IEEE, 2011, pp. 1–6.

[4] T. Kim, W. Wang, H. Tang, and K. Livescu, “Signer-independent fingerspelling recognition with deep neural network adaptation,” in Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, 2016, pp. 6160–6164.

[5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1. IEEE, 2005, pp. 886–893.

[6] T. Kim, J. Keane, W. Wang, H. Tang, J. Riggle, G. Shakhnarovich, D. Brentari, and K. Livescu, “Lexicon-free fingerspelling recognition from video: Data, models, and signer adaptation,” Computer Speech & Language, vol. 46, pp. 209–232, 2017.

[7] N. Pugeault and R. Bowden, “Spelling it out: Real-time asl fingerspelling recognition,” in Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, 2011, pp. 1114–1119.

[8] B. Kang, S. Tripathi, and T. Q. Nguyen, “Real-time sign language fingerspelling recognition using convolutional neural networks from depth map,” in Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on. IEEE, 2015, pp. 136–140.

[9] 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.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., “Imagenet large scale visual recognition challenge,” International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.

[11] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[14] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, “Densely connected convolutional networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, vol. 1, no. 2, 2017, p. 3.

[15] B. A. Dabwan, F. A. Olayah, H. T. Halawani, A. M. Mashraqi, Y. A. Abdelrahman and A. F. Shamsan, "Classification of Sign language Using VGG16 and VGG19," 2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS), Istanbul, Turkiye, 2023, pp. 1-6.

[16] Kaggle. (n.d.). "Sign Language MNIST." Retrieved August 26, 2022, from https://www.kaggle.com/datasets/datamunge/sign-language-mnist

[17] R. M. Badiger, R. Yakkundimath, and N. Malvade, "Deep Learning Based Classification of Double-Hand South Indian Sign Language Gestures for Deaf and Dumb Community," *Eur. Chem.,* vol. 12, no. 6, pp. 3191–3201, 2023.

[18] B. A. Dabwan, F. A. Olayah, H. T. Halawani, A. M. Mashraqi, Y. A. Abdelrahman and A. F. Shamsan, "Classification of Sign language Using VGG16 and VGG19," 2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS), Istanbul, Turkiye, 2023, pp. 1-6.

[19] K. Hong, G. Kim, and E. Kim, "GhostNeXt: Rethinking Module Configurations for Efficient Model Design," Applied Sciences, vol. 13, p. 3301.

[20] K. Shaheed, Q. Abbas, A. Hussain, and I. Qureshi, "Optimized Xception Learning Model and XgBoost Classifier for Detection of Multiclass Chest Disease from X-ray Images," Diagnostics, vol. 13, p. 2583, 2023.

[21]M. Shafiq and Z. Gu, "Deep Residual Learning for Image Recognition: A Survey," Applied Sciences, 2022, doi: 10.3390/app12188972.

[22] S. Tammina, "Transfer Learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images," International Journal of Scientific and Research Publications (IJSRP), vol. 9, p. 9420, 2019.

[23] R. Daroya, D. Peralta and P. Naval, "Alphabet Sign Language Image Classification Using Deep Learning," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 0646-0650.

[24] A. Leandro, C. Carneiro, “ A step further in the creation of a signlanguage translation system based on artificial intelligence“, TowardsData Science. (n.d.). https://towardsdatascience.com/a-step-further-in-the-creation-of-a-sign-language-translation-system-based-on-artificial-9805c2ae0562, (Accessed: August 26, 2022)