Research Article

Computer Vision-Enabled Character Recognition of Hand Gestures for Patients with Hearing and Speaking Disability

Sapna Juneja,[1] **Abhinav Juneja,**[1] **Gaurav Dhiman** ⓘ**,**[2] **Shashank Jain,**[3] **Anu Dhankhar,**[4] **and Sandeep Kautish** ⓘ[5]

[1]*KIET Group of Institutions, Delhi NCR, Ghaziabad, India*
[2]*Government Bikram College of Commerce, Patiala, India*
[3]*BMIET, Sonepat, India*
[4]*KIET Group of Institutions, Ghaziabad, India*
[5]*LBEF Campus, Kathmandu, Nepal*

Correspondence should be addressed to Sandeep Kautish; dr.skautish@gmail.com

Hand gesture recognition is one of the most sought technologies in the field of machine learning and computer vision. There has been an unprecedented demand for applications through which one can detect the hand signs for deaf people and people who use sign language to communicate, thereby detecting hand signs and correspondingly predicting the next word or recommending the word that may be most appropriate, followed by producing the word that the deaf people and people who use sign language to communicate want to say. This article presents an approach to develop such a system by that we can determine the most appropriate character from the sign that is being shown by the user or the person to the system. To enable pattern recognition, various machine learning techniques have been explored and we have used the CNN networks as a reliable solution in our context. The creation of such a system involves several convolution layers through which features have been captured layer by layer. The gathered features from the image are further used for training the model. The trained model efficiently predicts the most appropriate character in response to the sign exposed to the model. Thereafter, the predicted character is used to predict further words from it according to the recommendation system used in this case. The proposed system attains a prediction accuracy of 91.07%.

1. Introduction

Machine learning [1] has changed the dimensions of how society thinks and perceives the world. There is a multidimensional growth in the range of applications [2] that are being supported by machine learning. All the three formats of machine learning including supervised, unsupervised, and reinforcement learning have been finding their solicitations in newer ways, affecting human lives and giving an all-new shape to how modern-day businesses are carried out. Among the various application of machine learning, computer vision is one area that is attracting the research community to harness the potential of technological developments for felicitating society. Computer vision [3] has been the field of active research and

development with the developments in the field of computer vision technology. Researchers are employing computer vision for the classification of static images [4], and a plethora of algorithms have been proposed for the same. Similarly, computer vision is playing a vital role in medical sensing and diagnostics [5]. There is an inherent need for such a system in society to enable deaf people and people who use sign language to communicate in their day-to-day activities. With computer vision, our machines can be empowered to recognize their hand gestures. "What do they want to convey by their gestures?" If known, anyone can interpret their needs through those gestures. In this article, we have designed a robust system, which is capable of predicting the words with the help of hand signs. The hand sign denotes what a particular sign means.

Thereby, after determining the characters further recommending them the best possible word, they can choose from to make the system hassle free and easy to use so that they can use this system more accurately and precisely according to their needs. The general approach involves the detection of signs with the help of a convolutional neural network (CNN) through which a model has been created [6]. The model can extract features through its several layers; every layer, by recognizing certain features, determines whether that particular new image matches the training dataset. With this approach, we propose an effective and intelligent hand gesture recognition [7] application.

2. Literature Review

Machine learning is not a new technology that has recently evolved. There has been a quest to make the world better using artificial intelligence and machine learning since the 1970s. Wang et al. [8] exhibited turn invariant stances utilizing limit histogram [9, 10]. A camera has been used to secure the information picture. The channel for skin shading discovery [11] is being utilized by the bunching procedure to discover the limit for each gathering in the grouped picture, utilizing a common form following calculation. The picture was divided into multiple networks, and the limits were normalized. The limit was referred to as harmony's size chain which was utilized in the form of a histogram, by separating the picture into a number of areas N in an outspread structure, according to the explicit edge. For the classification process, neural networks [12], MLP and dynamic programming, and DP coordinating were utilized. Numerous analyses have been executed on various highlight positions notwithstanding utilizing distinctive harmony's size histogram and harmony's size FFT [13]. Convolution neural networks [14, 15] have an established role in image recognition as has been evidentially demonstrated by several researchers in past. The specific contribution of CNNs to medical disease diagnosis with the help of relating the scanned images to the presence or absence of diseases [16] is an exceptional application with proven efficiency and reliability. Rectified Linear Unit (ReLU) [17] is one of the most robust activation functions that is employed for image processing. ReLU is one of the most popular nonlinear activation functions trusted by researchers for deep learning projects. 26 static stances from American sign language were utilized in the trials. A homogeneous foundation was applied in the work. Stergiopoulou proposed another self-growing and self-organized neural gas (SONG) network [18] for hand motion acknowledgment. For hand district discovery, a shading division procedure dependent on the skin shading channel in the YCbCr shading space [19] was utilized, and an estimation of hand-formed morphology has been recognized utilizing SONG organize; three highlights were extricated, utilizing the finger-distinguishing proof procedure that decides the quantity of the lifted fingers and quality of shape of the hand, and the Gaussian circulation model was utilized for acknowledgment. Table 1 shows a review of the contributions by the researchers in a similar domain as the proposed work.

3. Proposed Methodology

The current work has been summarized graphically with the help of Figure 1. The implementation involves categorizing the classes that the person using this application wants to predict. The task is to recognize the hand gestures and classify them as particular tasks which are basically the predicted values that are to be categorized. Figure 1 majorly showcases the concept of the pipeline that has been employed to train the deep neural network [30] (Gupta et al., 2021). The first convolution neural network weights are trained with certain feature values being modified; i.e., certain feature values are extracted from the images of the dataset. The model is being trained upon; then, after passing on all the images to the model, it updates its weight and all other parametric values [31, 32] on which the model relies [33].

Then, to establish trust and reliability on the developed application, we have used a testing dataset. After performing several testing and tuning initiatives, the best and most accurate model is taken into consideration. The process of optimization is reiterated; our model is trained and tested upon again and again to find out the best model for considerations [34].

3.1. Datasets. Initially, for training the proposed model, we prepared our custom dataset, which is related to American sign language and consists of the combination of different signs and their corresponding labels. Accordingly, the model has been trained using the same dataset. We have taken further the MNIST image dataset from open repository Kaggle.

3.1.1. CSV Files for Performing Analysis. The snapshot of the training CSV file (Figure 2) that consists of a variety of hands symbols and their given pixel values is used from which we have created the script to extract the images by using the np. array function and then generate the respected images from it. Similarly, for test CSV, images have been created using the pixel values from the given set of CSV files and generating the respective images and then storing them in the testing folder.

3.1.2. Training Images. As there is a need for the creation of a reliable model to perform its corresponding best prediction, we have created a customized dataset for the process. This involves certain steps such as binary image creation [35], background removal [36], and edge detection for the creation of that particular dataset [37]. This was carried out to ensure that the responses are more appropriate, though this may be generalized to any similar gesture for a similar type of response from the system.

The dataset employed for modeling consists of 23,826 images in a total of $28 * 28$ pixel images, each with different labels for each set of images for each character from A–Z (Figure 3).

TABLE 1: Brief review of the work carried out on gesture recognition by various researchers.

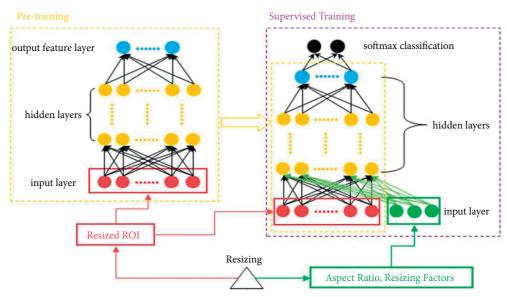| S. no.	Author, year	Technique	Objective and outcome
1	Tran et al., 2020 [28]	3 DCNN	Fingertip detection with hand gesture recognition was achieved with good accuracy levels for recognition of the patterns on a big dataset
2	Huang et al., 2019 [22]	Skin detection and deep learning	Firstly detects the skin followed by contour and segment recognition and finally gesture identification for various 10 symbols
3	Fhager et al., 2019 [21]	Pulsed radar and CNN	Framed time envelope scheme is used to represent data, and further machine learning operations are performed to extract gesture information
4	Aly et al., 2019 [20]	Deep learning and PCA net	The proposed model used principal component analysis for feature selection and SVM for classifying the gestures
5	Sanchez-Riera et al., 2018 [27]	ICP minimization and deep learning	This article introduces the variability adjustment of hand gestures of different persons with a high degree of accuracy
6	Lai et al., 2018 [24]	DSP using FSP algorithm	In this work, a real-time hand gesture evaluation system has been proposed. It is also explored for use in volume control and TV channel changing. The work had an accuracy of 94.3%
7	Ryu et al., 2018 [26]	FCFW and feature analysis	Gesture recognition is performed by FCFW, and a new feature extraction technique is proposed which performs fairly well compared to the existing techniques
8	Kim et al., 2017 [23]	Ultra-wideband impulse and CNN	Hand gesture cognition is performed using ultra-wideband impulse sensors in this work. The model predicts 6 hand gestures with 90% accuracy
9	Chen et al., 2014 [29]	Background subtraction technique	Used the background subtraction method for gesture recognition on a dataset of 1300 images
10	Lee and Tanaka, 2013 [25]	Depth-based hand recognition	Evaluated 6 hand gestures with volunteers and achieved an accuracy of nearly 91% for the recognition of gestures

FIGURE 1: Flow chart for the proposed system.

3.1.3. Testing Images. For testing the model accuracy, we need a certain set of images with their corresponding labels for the model so that it can predict the accurate results for the creation of a better model (Figure 4).

This dataset consists of 2,668 images with their corresponding labeled folders and with the corresponding images inside it.

3.2. Data Preprocessing Phase [38]. This phase is most instrumental in grabbing all the necessary data for predicting. If required, this phase involves cleaning of data that is

needed for feeding the machine with the most prescribed data to perform the best analysis on it. Certain common steps that were resolved to perform data analysis include the blurring of the background and edge detection, which are performed on the image to determine the hands. This involves the mapping of certain pixels to confirm the best-required data for performing specific analysis on the data.

This process is important as all the unwanted information is removed from the image in order to perform better training of the model. Furthermore, after the creation of a specific data matrix of 28 ∗ 28 pixels passing
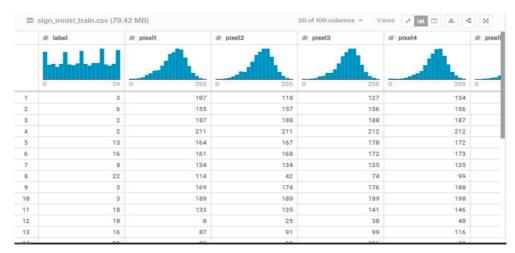
Figure 2: Training dataset CSV file.

on to the model, the model is trained on the required information that is needed to train on the images of that mapping space.

3.3. Algorithms Used.
Convolution neural networks (Dhiman et al., 2021) have been preferred as the neural networks [39] used for the extraction of the most certain features in an image (Figure 5). This is the type of neural network by which certain pixel information is extracted from the image and then stored in certain layers to perform the analysis. CNN is used for the extraction of certain key features through which one can recognize the image. This involves certain training phase information by which one can extract the information from the images of certain labels and then mapping that information to those labels and generating the most requisite information that is needed to predict the images [40].

A convolution neural network involves certain layers such as the convolution layer, pooling layer, and the ANN structure. The convolution layer is the layer that is responsible for extracting the features which involves detection of edges, corners, etc. The extraction of edges from an image is followed by training the weights along with saving those edges over that neural network which are used to determine the particular label over the testing phase. So, after determining those values with the help of the model, we can easily predict the labels used for predicting certain images over that local space.

Contributing layers that are used for extraction of features over the space are as follows.

3.3.1. Convolution Layer [41].
Convolution is a dedicated linear operation employed for the purpose of extraction of significant features. A kernel, which is typically a small number array, is applied at the input and named a tensor. Product across every element belonging to the kernel and the input tensor is generated for every point of the tensor, and for obtaining the output values, all these are summed up. This is also referred to as the feature map. This process is recursively carried out to generate a huge number of feature

Figure 3: Training image.

Figure 4: Testing image.

maps by using multiple kernels. With this, the different kernels may be interpreted as the feature extracting agents. The layer is instrumental for the extraction of edges from the images and storing those over the network to form those features to map them over some required neurons.

3.3.2. Batch Normalization [42].
Batch Normalization is used to increase the stability of the neural network with the reduction in the covariance shift. It normalizes the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation. The typical role of a batch normalization layer includes enabling every layer to learn on its own, independent of the other network layers.

3.3.3. ReLU Activation [43].
The rectified linear activation function is a piecewise linear function that will output the input directly if it is positive; otherwise, it will output zero. It has become the default activation function for many types of neural networks because a model that uses it is easier to train
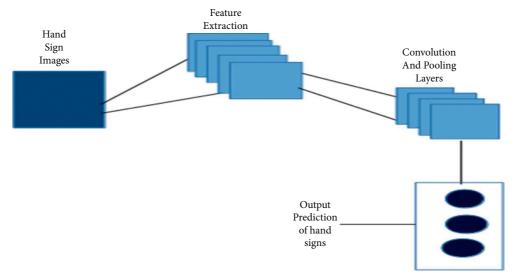
FIGURE 5: A convolution neural network.

and often achieves better performance. This is used generally in the first layers which provides ease of evaluating the output to the next neural network layers. The ReLU function may be described using equations (1) and (2). ReLU has a constant gradient for the positive input.

$$\text{ReLu}(x) = \text{Max}(0, x), \tag{1}$$

$$\frac{d}{dx}\text{ReLu}(x) = \{1, \text{ if } x > 0; \text{ otherwise}\}. \tag{2}$$

3.3.4. Zero Padding 2D [44]. This involves adding rows or columns of the zeroth layer either to the top, bottom, right, or left to the neural network. Padding is carried out to retain the size of the tensor as it is lost in the convolution layer. It retains the in-plane dimensions of the frame; otherwise, each successive frame would get smaller in size.

3.3.5. Max Pooling 2D [45]. The pooling layer plays the role of reducing the in-plane dimensionality of the feature map with the help of downsampling. The max pooling operation involves taking a certain kernel of a certain size, then finding out the maximum value from the kernel, and then by extracting the pixel with the higher value, sending them to the next layer. There may be changes in the height features of the feature map, but the depth feature remains unaltered.

3.3.6. Softmax Activation Function [46]. The softmax function is a wonderful activation function that turns numbers aka logits into probabilities that sum to one. The softmax function outputs a vector that represents the probability distributions of a list of potential outcomes.

4. Implementation

This project work has been executed using a layered approach. To perform this research work, we have created 5 layers including one fully connected layer to perform the analysis of the data. The block diagram given in Figure 6 shows the flow of our experimental process. The first step, being the data preprocessing step, cleans the data to find more accurate data to analyze the information. The cleaning of data is necessary to have accuracy and reliability in the estimation of the correct response. Various data cleaning methodologies [47] have been employed for this purpose. After that, the data are forwarded to the next layer for training. The training layer involves the extraction of the features which are of more significance for decision making. Training is performed based on those selected features. For the current work, we have used the convolution layer which involves 32, 64, 128, and 256 sizes of those layers, and then, the convolved image is passed forward to perform batch normalization [48]. Batch normalization provides fast and accurate processing. It is further passed to the RelU activation function to categorize the neurons to perform the training, which is further passed to the max pooling layer to find out the maximum pixel that could contribute to the training of an image. After that, it is passed to a fully connected dense layer of the network to analyze the results using the softmax function.

After the training phase, the model is trained on a variety of sets of images, for the signs that we need to recognize those gestures made by the person in the live feed. Therefore, we need to pass images as per the frame to the model. Subsequently, the model recognizes each and every frame passed to it and will recognize the area allotted to it or passed inside it; this will be the region of interest (ROI) [49]. In our case, we take the ROI area and convert that multiple sets of images into the array of pixel values, and the array of pixel
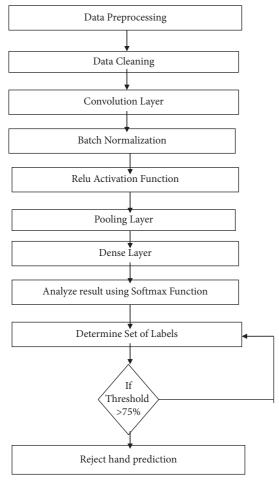
Figure 6: Block diagram.

values is converted into floating values to be transferred to them. After retrieving the floating-point values of the set of images being transferred to them based on its intensity values, the model is trained for identification of those particular regions with certain probabilities; i.e., it will recognize every character of alphabets with a certain value of probability.

The proposed model has been trained on 100 epochs and keeping a batch size of 32 to perform better and more accurate predictions on the data. Similarly, for testing out the results, it was anticipated to pass the image on to the neural network and predict certain results based on the trained model. Subsequently, the results are generated as the label of data or the word as the output that any person is portraying by his hand. Generalizing it further, it involves the creation of a model for better prediction of results. For achieving this goal, a trained model on the desired set of $28 * 28$ px images is generated using the hand sign MNIST dataset [50]. In our work, we have been able to create those images and then create a framework on which predictions can be made for hand gestures experienced by the system. In order to ensure that the results are not skewed, CNN was employed which involves extraction of the images by performing edge detection [51], noise removal [52, 53], and corner detection [30] for mapping those results using those layers. Figures 7

and 8 show the sample region of interest images from the used dataset. For evaluation and real-time prediction of results, the model is needed to be passed with certain required data to be predicted, and for that, an appropriate frame has been created, so that analysis will be only taken on the frame.

By using this, image boundaries are analyzed and cropped images are formed for the better analysis of the model to produce accurate results using the model. This way, our model is trained on these images, and for the real-time object detection, the frame is taken to produce those results as shown in Figure 9 and to analyze those images for better prediction of results for the real-time detection.

This frame is initially captured, and the analysis is made on this frame to extract the maximum features of the gestures [54]. The information is assimilated accordingly, and based on this, the predictions are made. Then, the applied image model returns a certain probability score, and using np.argmax [55], the maximum probability output is selected and the corresponding result is displayed on the screen of the character which is needed to get an output response.

5. Result Analysis

In the proposed work, using Opencv [56], library cv2.putText, we have proposed a model to predict the most probable characters using minimum threshold selection. We had taken this threshold as 95% in the current case to represent those values, which determines the outcome of the model to be provided as the predicted gestures. Overall, we captured the best probable sign that is predicted by the model. Figures 10 and 11 show the typical response of our system for the hand gestures that have been given as inputs to the system.

The proposed system works with an accuracy of 91.07% in its prediction for the gestures. To achieve this accuracy, we have created 2 separate folders, based on training and test set size which is obtained by creating the custom dataset having 23,826 training and 2,668 test set images of different labeled folders. Furthermore, we specified those characters where we have used CNN models to train on those sets of images. This was trained with 100 epochs with a batch size of 32 to obtain that accuracy for those images which have been augmented with hand images by using certain convolution and other linear operations to increase the set of images.

As those images are $28 * 28$ pixels, we had 05 layers followed by a fully connected neural network layer in which all five layers consist of one convolution layer which extracts convolved features from the images; then, those features are passed to the batch normalization layer to smoothen the images. Later, they are passed to the max pooled layer which will grab those higher-valued features that mostly affect our images, which are finally passed to the activation function. This ReLu in our case is used to obtain a certain set of weights on which the model is being trained upon to grab certain weights.

Similarly, the model is being trained with 100 epochs which represents 100 times the weights are updated to follow-up the prediction of the model. Thereby, after

FIGURE 7: ROI image.

FIGURE 8: ROI image.
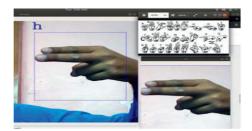

FIGURE 9: Real-time feed region of interest.

FIGURE 10: Sample 1: real-time gesture prediction.

retraining on those images, we obtain 91.07%, as shown in Figure 12. The architecture in our case is passed to a certain ROI region, meaning, thereby, any hand within that area will be predicted with 91.07% accuracy, leading our model to be away from any distortions that came outside our predicting area which is the ROI area in our case.

We have been able to identify all the 26 characters of English for the American sign language with the above-mentioned accuracy levels. The articles which we used as the basis of our work were a source of inspiration for the current work and have been the foundation for the basic knowledge and understanding of the experimental work performed. Table 2 shows the comparative analysis of the proposed system with respect to the existing work of relevant work as the current work.

As can be observed from Table 2 that generally, most of the proposed models in the earlier works have focused on a limited number of gestures, in contrast, the proposed work has considered a comprehensive set of all the characters to be mapped to a corresponding alphabet.

FIGURE 11: Sample 2: real-time gesture prediction.

FIGURE 12: Result evaluation.

TABLE 2: Outcomes of some of the earlier techniques and current work.

S. no.	Author, year	Efficiency (%)	Remarks
1	Kim et al., 2017 [23]	90	Depicts 6 hand gestures only
2	Lee and Tanaka, 2013 [25]	91	Depicts 6 hand gestures
3	Current work	91.07	Depicts 26 hand gestures and comprehensively predicts the different samples

6. Conclusions

With the help of training our deep learning [57] convolution neural networks, it is simply an evolutionary step involved in generalizing a model by which we can help deaf people and people who use sign language to communicate. This current research work is an exemplary work that revolves around the analysis by which we had created the model through which our machine will be able to recognize what we want to communicate with them.

So, with the help of the CNN, our machine will be able to identify the patterns in our data, and with the help of those patterns, our CNN keeps track of the weights and other parameters by which we can identify our machine so that it will be easier for our machine to keep track of certain parameters by which our model can identify the patterns, and based on certain parameters such as weights and the other convolution parameters used while building the CNN in our convolution layers, we can identify those images on which our model is trained upon and after training our layers by itself determining certain convolution matrix [58].

To identify the convolution patterns to express those layers in our neural network, after that, we tested our model by using real-time video frames and then giving certain ROIs on which we are training our neural network to identify those patterns in our data, and then, our model predicted certain results through which any person can communicate with each other by using certain hand gestures and can transmit the information from one user and can communicate with any person easily with the help of computer vision and deep learning media to transmit the information.

The model worked up to the expectations of the objectives of the current research work. During the course of action, it was found that the model suffers a few limitations. The first one is variation in the performance of the DL model with different lighting conditions such as brightness and intensity. The second observation is that the model might fail to predict some hand gestures of movement such as it can predict only stable hand gestures. For deaf people and people who use sign language to communicate, some moving hand gestures represent particular characters, and the current work may not be able to accurately predict some of those.

The current work may further be explored and enhanced to make it more effective and reliable for a physically challenged person by creating a chat-based system so that deaf people and people who use sign language to communicate can chat through the system by just using their hand gestures. They can create words from the system and send them to others instantly without the need to actually type the text.

Data Availability

The data used to support the findings of this study are available from the author upon request (dr.skautish@gmail.com).

Ethical Approval

This study was approved by the Department of Biotechnology, IMS Engineering College, Ghaziabad, India.

Consent

Appropriate consent was taken from the participants for enrolment.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Sapna Juneja performed conceptualization and formal analysis, developed the methodology, and wrote the initial draft. Abhinav Juneja wrote the original draft and performed supervision. Gaurav Dhiman performed visualization and performed project administration. Shashank Jain performed formal analysis, and Anu Dhankhar reviewed and edited the article. Sandeep Kautish wrote the final draft of the paper.

References

[1] Z. Gong, P. Zhong, and W. Hu, "Diversity in machine learning," *IEEE Access*, vol. 7, pp. 64323–64350, 2019.

[2] R. Sharma and K. Davuluri, "Design patterns for machine learning applications," in *Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2019.

[3] V. Wiley and T. Lucas, "Computer vision and image processing: a paper review," *International Journal of Artificial Intelligence Reserach*, vol. 2, no. 1, pp. 28–36, 2018.

[4] M. I. Lakhal, H. Çevikalp, S. Escalera, and F. Ofli, "Recurrent neural networks for remote sensing image classification," *IET Computer Vision*, vol. 12, no. 7, pp. 1040–1045, 2018.

[5] J. Thevenot, M. B. Lopez, and A. Hadid, "A survey on computer vision for assistive medical diagnosis from faces," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1497–1511, 2018.

[6] A. Juneja, S. Juneja, A. Soneja, and S. Jain, "Real time object detection using CNN based single shot detector model," *Journal of Information Technology Management*, vol. 13, no. 1, pp. 62–80, 2021.

[7] V. S. Kulkarni and S. D. Lokhande, "Appearance based recognition of American sign language using gesture segmentation," *International Journal on Computer Science and Engineering IJCSE*, vol. 2, no. 3, pp. 560–565, 2010.

[8] Z. Wang, S. Wu, C. Liu, and S. Wu, "The regression of MNIST dataset based on convolutional neural network," in *Proceedings of the 2020 International Conference on Advanced Machine Learning Technologies and Applications*, Jaipur, India, 2020.

[9] A. Thakur and D. S. Saini, "Correlation processor-based sidelobe suppression for polyphase codes in radar systems," *Wireless Personal Communications*, vol. 115, 2020.

[10] S. Zhao, W. Tan, S. Wen, and Y. Liu, "An Improved algorithm of hand gesture recognition under intrikate pond," *International Conference on Intelligent Robotics and Applications*, pp. 786-787, 2008.

[11] J. Sancho Rieger, S. Waldert, T. Pistohl et al., "Vision based gesture recognition using neural networks approaches: a review," *International Journal of Human Computer Interaction*, vol. 32, no. 1, pp. 480–494, 2008.

[12] M. Hagen, A. E. Ayadi, J. Wang, N. Vasiloglou, and E. Afshar, "Optimizing training data for image classifiers," in *Proceedings of the KDD 2019 Workshop on Data Collection, Curation, and Labeling for Mining and Learning*, 2019.

[13] B. Shubankar, M. Chowdhary, and M. Priyaadharshini, "IoT device for disabled people IoT device for disabled people," *Procedia Computer Science*, vol. 165, pp. 189–195, 2020.

[14] J. Gu, Z. Wang, J. Kuen et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, 2006.

[15] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," *Procedia Computer Science*, vol. 132, pp. 377–384, 2018.

[16] R. H. Abiyev and M. K. S. Ma'aitah, "Deep convolutional neural networks for chest diseases detection," *Journal of Healthcare Engineering*, vol. 2018, Article ID 4168538, 11 pages, 2018.

[17] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, https://arxiv.org/abs/1803.08375.

[18] M. Yasen and S. Jusoh, "A systematic review on hand gesture recognition techniques, challenges and applications," *PeerJ Computer Science*, vol. 20, 2019.

[19] J. Liu, K. Furusawa, T. Tateyama, Y. Iwamoto, and Y. Chen, *An Improved Hand Gesture Recognition with Two-Stage Convolution Neural Networks Using a Hand Color Image and its Pseudo-Depth Image Graduate School of Information Science and Engineering*, Ritsumeikan University, Kyoto, Japan, 2019.

[20] W. Aly, S. Aly, and S. Almotairi, "User-independent american sign language alphabet recognition based on depth image and PCANet features," *IEEE Access*, vol. 7, pp. 123138–123150, 2019.

[21] L. O. Fhager, S. Heunisch, H. Dahlberg, A. Evertsson, and L. E. Wernersson, "Pulsed millimeter wave radar for hand gesture sensing and classification," *IEEE Sensors Letters*, vol. 3, no. 12, p. 1, 2019.

[22] H. Huang, Y. Chong, C. Nie, and S. Pan, "Hand gesture recognition with skin detection and deep learning method," *Journal of Physics: Conference Series*, vol. 1213, no. 2, 2019.

[23] S. Y. Kim, H. G. Han, J. W. Kim, S. Lee, and T. W. Kim, "A hand gesture recognition sensor using reflected impulses," *IEEE Sensors Journal*, vol. 17, no. 10, pp. 2975-2976, 2017.

[24] H. Y. Lai, H. Y. Ke, and Y. C. Hsu, "Real-time hand gesture recognition system and application," *Sensors and Materials*, vol. 30, no. 4, pp. 869–884, 2018.

[25] U. Lee and J. Tanaka, "Finger identification and hand gesture recognition techniques for natural user interface," in *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction*, Bangalore, India, 2013.

[26] S. J. Ryu, J. S. Suh, S. H. Baek, S. Hong, and J. H. Kim, "Feature-based hand gesture recognition using an FMCW radar and its temporal feature analysis," *IEEE Sensors Journal*, vol. 18, no. 18, pp. 7593–7602, 2018.

[27] J. Sanchez-Riera, K. Srinivasan, K. L. Hua, W. H. Cheng, M. A. Hossain, and M. F. Alhamid, "Robust RGB-D hand tracking using deep learning priors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2289–2301, 2018.

[28] D. S. Tran, N. H. Ho, H. J. Yang, E. T. Baek, S. H. Kim, and G. Lee, "Real-time hand gesture spotting and recognition

using RGB-D Camera and 3D convolutional neural network," *Applied Sciences (Switzerland)*, vol. 10, no. 2, 2020.

[29] Z.-h. Chen, J.-T. Kim, J. Liang, J. Zhang, and Y. B. Yuan, "Real-time hand gesture recognition using finger segmentation," *The Scientific World Journal*, vol. 2014, Article ID 267872, 9 pages, 2014.

[30] K. Martin Sagayam, S. Suresh, D. Jude Hemanth, L. Henessey, and C. Ho, "Optimization of SVM-based hand gesture recognition system using particle swarm optimization and plant growth simulation algorithm," *The Biometric Computing*, CRC Press, Boca Raton, FL, USA, 2019.

[31] I. G. Putra and K. Negara, "A survey of hand gesture recognition methods in sign language recognition," *Petranika Journal of Science & Technology*, vol. 26, 2018.

[32] N. Yuvaraj, V. Chang, B. Gobinathan et al., "Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification," *Computers and Electrical Engineering*, vol. 92, Article ID 107186, 2021.

[33] S. Juneja, M. Gahlan, G. Dhiman, and S. Kautish, "Futuristic cyber-twin architecture for 6G technology to support internet of everything," *Scientific Programming*, vol. 2021, Article ID 9101782, 7 pages, 2021.

[34] S. Juneja, A. Juneja, G. Dhiman, S. Behl, and S. Kautish, "An approach for thoracic syndrome classification with convolutional neural networks," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 3900254, 10 pages, 2021.

[35] P. Anjankar and S. Borkar, "Artificial intelligence based robot control using face and hand gesture artificial intelligence based robot control using face and hand gesture recognition: a review," 2015, https://www.researchgate.net/publication/326008958_Artificial_Intelligence_Based_Robot_Control_Using_Face_and_Hand_Gesture_Recognition_A_Review.

[36] Q. Wang and Y. Yuan, "Learning to resize image," *Neurocomputing*, vol. 131, pp. 357–367, 2014.

[37] A. Sharma and R. Kumar, "Computation of the reliable and quickest data path for healthcare services by using service-level agreements and energy constraints," *Arabian Journal for Science and Engineering*, vol. 44, no. 11, pp. 9087–9104, 2019.

[38] K. K. Al-jabery, T. Obafemi-Ajayi, G. R. Olbricht, and D. C. Wunsch, "Data preprocessing," *Computational Learning Approaches to Data Analytics in Biomedical Applications*, pp. 7–27, 2020, in press.

[39] M. H. Sazli, "A brief review of feed-forward neural networks," *Communications*, Faculty of Science, University of Ankara, Ankara, Turkey, 2006.

[40] S. K. Korde and K. Jondhale, "Hand gesture recognition system using standard fuzzy C-means algorithm for recognizing hand gesture with angle variations for unsupervised users," in *Proceedings of the 2008 1st International Conference on Emerging Trends in Engineering and Technology*, Nagpur, India, 2008.

[41] H. Sedghi, V. Gupta, and P. M. Long, "The singular values of convolutional layers," 2019, https://arxiv.org/abs/1805.10408.

[42] Z. Chen, L. Deng, G. Li et al., *Batch Normalization Sampling*, 2019.

[43] A. Petrosyan, A. Dereventsov, and C. G. Webster, "Neural network integral representations with the ReLU activation function," 2019, https://arxiv.org/abs/1910.02743.

[44] I. Domuta and T. P. Palade, "Sliding DFT and zero padding," in *Proceedings of the 2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, Budapest, Hungary, 2019.

[45] V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Kr, and A. Maier, "Deep generalized max pooling," 2020, https://arxiv.org/abs/1908.05040.

[46] I. Kouretas and V. Paliouras, "Simplified hardware implementation of the softmax activation function," in *Proceedings of the 2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, Thessaloniki, Greece, 2019.

[47] Y. Wang, Z. Pan, and Y. Pan, "A training data set cleaning method by classification ability ranking for the k-nearest neighbor classifier," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, pp. 1544–1556, 2019.

[48] Z. Dai and R. Heckel, "Channel normalization in convolutional neural network avoids vanishing gradients," 2019, https://arxiv.org/abs/1907.09539.

[49] M. Haindl, "Region of interest contrast measures," *Kybernetika -Praha-*, vol. 54, no. 5, pp. 978–990, 2018.

[50] D. Bragg, O. Koller, M. Bellard et al., "Sign language recognition, generation , and translation: an interdisciplinary perspective," in *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, Pittsburgh, PA, USA, 2019.

[51] K. O. Jimoh, A. O. Ajayi, and I. K. Ogundoyin, "Template matching based sign language recognition system for android devices," *FUOYE Journal of Engineering and Technology*, vol. 5, 2020.

[52] R. Z. Khan and N. Ibraheem, "Hand gesture recognition: a literature," *International Journal of Artificial Intelligence & Applications*, vol. 3, no. 4, pp. 161–174, 2012.

[53] A. Thakur and D. S. Saini, "Bandwidth optimization and sidelobe levels reduction in PC radar using Legendre orthogonal polynomials," *Digital Signal Processing: A Review Journal*, vol. 101, Article ID 102705, 2020.

[54] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: a review," *Computer Vision and Image Understanding*, vol. 108, 2007.

[55] P. Kirci and M. F. Cambek, "Hand gesture detection," in *Proceedings of the 2019 International Conference on Information Technologies (InfoTech)*, Varna, Bulgaria, 2019.

[56] Y. Guan and M. Zheng, "Real-time 3D pointing gesture recognition for natural HCI," in *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, Chongqing, China, 2008.

[57] V. Ganesan, "Computer aided system for autism spectrum disorder using deep learning methods," *International Journal of Psychosocial Rehabilitation*, 2020, in press.

[58] W. T. Freeman and C. D. Weissman, "Television control by hand gestures," in *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, 1995.