Recognition of Hand Gesture Image Using Deep Convolutional Neural Network

Article *in* International Journal of Image and Graphics · January 2021

DOI: 10.1142/S0219467821400088

Martin Sagayam
Karunya University

A. Diana Andrushia
Karunya University

Ahona Ghosh
Maulana Abul Kalam Azad University of Technology West Bengal

Omer Deperlioglu
Afyon Kocatepe University

International Journal of Image and Graphics
Vol. 21, No. 1 (2021) 2140008 (15 pages)

DOI: 10.1142/S0219467821400088

Recognition of Hand Gesture Image Using Deep Convolutional Neural Network

K. Martin Sagayam[*,¶], A. Diana Andrushia[*,‖], Ahona Ghosh[†,**],
Omer Deperlioglu[‡,††] and Ahmed A. Elngar[§,‡‡]

[*]*Department of Electronics and Communication Engineering*
Kandiarunya Institute of Technology and Sciences
Coimbatore 641114, India

[†]*Department of Computational Science, Brainware University*
Kolkata 700125, India

[‡]*Department of Computer Programming*
Afyon Kocatepe University
Afyonkarahisar, Turkey

[§]*Faculty of Computers and Artificial Intelligence*
Beni-Suef University, Beni-Suef, 62511, Egypt
[¶]*martinsagayam.k@gmail.com*
[‖]*andrushia@gmail.com*
[**]*ahg.cs@brainwareuniversity.ac.in*
[††]*deperlioglu@gmail.com*
[‡‡]*elngar_7@yahoo.co.uk*

Received 23 November 2020
Accepted 23 November 2020
Published 29 January 2021

In recent technology, there is tremendous growth in computer applications that highlight human–computer interaction (HCI), such as augmented reality (AR), and Internet of Things (IoT). As a consequence, hand gesture recognition was highlighted as a very up-to-date research area in computer vision. The body language is a vital method to communicate between people, as well as emphasis on voice messages, or as a complete message on its own. Thus, automatic hand gestures recognition systems can be used to increase human–computer interaction. Therefore, many approaches for hand gesture recognition systems have been designed. However, most of these methods include hybrid processes such as image pre-processing, segmentation, and classification. This paper describes how to create hand gesture model easily and quickly with a well-tuned deep convolutional neural network. Experiments were performed using the Cambridge Hand Gesture data set for illustration of success and efficiency of the convolutional neural network. The accuracy was achieved as 96.66%, where sensitivity and specificity were found to be 85% and 98.12%, respectively, according to the average values obtained at the end of 20 times of operation. These results were compared with the existing works using the same dataset and it was found to have higher values than the hybrid methods.

Keywords: Hand gesture recognition; deep neural network; convolutional neural network; human–computer interaction.

[‡‡]Corresponding author.

1. Introduction

Hand movements are widely used in human nonverbal interaction. These allow expressing orders, mood or some basic cardinal information. It can also be the only way to communicate in specific situations, such as traffic organization in the absenteeism of traffic lights for hearing-impaired people and in the absence of traffic lights.[1] Hand movement recognition is one of the maximum investigated fields recently highlighted in computer vision and machine learning. The most important goal here is to develop effective and easy systems for human–computer interaction. Because hand movements are natural and strong acting as a means to communicate with people. It also allows more people to communicate without the need for any special equipment. It can also be used for the interpretation and knowledge of sign languages for the hearing impaired. However, hand movement recognition is a very complex issue, given the example that there are many different sign languages.[2,3]

One of the main aims of human–machine interaction is to provide efficient and variable interaction. For example, the visual hand gesture recognition (HGR) system can be used in remote contact-free interactive environments such as operating rooms and intensive care in hospitals. It can also provide attractive interactions for gaming and entertaining applications. However, the interaction with the HGR standard keyboard and mouse is not as efficient. Sensitivity to size and speed changes can perform unwell against complicated backgrounds. Varying lighting circumstances also affect sensitivity. The determination of the possible hand movement stage restricted the use of hand movements as an appropriate method in designing the interface.[4,5]

Therefore, in recent times, several hand gesture recognition methods have been planned. For example, Wang *et al.* explained a new HGR system built on the distance measure of a new canonical super pixel graphic displacement. This study's goal is to improve the performance of the distance of the super-pixel displacement, which is a newly designed distance measure applied to depth-based hand movement recognition.[6] Stergiopoulou and Papamarkos proposed an approach for HGR that is created on a hand gesture right process through a new Self-Growing and Self-Organized Neural Gas network. At first, the portion of the hand has been recognized in the YCbCr color space. Second, they extracted characteristics with Self-Growing and Self-Organized Neural Gas network. Finally, they accomplished the HGR with a likelihood-based classification technique.[2] In another study, motion recognition system was proposed which did not use any marker. This HGR system is intended to recognize movements that cover almost all aspects of HGR such as system functions, launching applications and opening some popular websites.[7] In their work, Nasser *et al.* identified a real-time system containing three units: hand detection and monitoring using facial removal, skin knowing and contour contrast algorithm, posture recognition using a multi- class SVM and feature bag and a plurality of motion commands by following the transitions between stops.[8]

Image classification is explained as the operation of distinguishing images from one of the different predefined classes and is a basic problem in computerized vision. In current years, it has been shown that deep learning models overcoming these problems have benefited from a large number of nonlinear computing layers for pattern analysis and classification, as well as for feature extraction and transformation.[9] Deep learning networks like Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are broadly used in the HGR systems.

Shin and Sung developed two dynamic HGR methods applying less complex (RNN algorithms. The first one is built on a video signal and uses the unified building of a CNN and a RNN. The latter uses only one RNN using the accelerometer data. In hardware and software-based applications, they have implemented static point optimization that measures most of the weights as two bits to reduce power consumption and optimize the memory size for storing the weight.[10] Dadashzadeh *et al.* proposed a two-stage HGR system using CNN and called HGR-Net. In this approach, the first stage is semantic partitioning at the pixel level towards the hand zone. The second stage describes the hand movement type. The segmentation stage architecture is created on a mixture of the CNN and the spatial pyramid pool of the atmosphere.[11] Molchanov *et al.* proposed an algorithm from the challenging depth and density data using 3D CNN for drivers' HGR. This method combines information for more than one spatial scale for the last estimate. It also uses spatiotemporal data enhancement to reduce more effective training and potential overcompliance.[12]

The use of multiple frames in HGR systems reduces the classification accuracy of the classifier as well as the computational complexity. John and his colleagues proposed a method to solve these problems by introducing fewer representative frameworks from the video series and entering them into long-term recurring CNNs. They proposed to use new tiled picture designs and a tiled pair pattern on CNN to remove representative frames.[13] Tang and others proposed a two-stage Kinect based HGR system for Sign Language Recognition. At first, they used an algorithm to design a system for detecting and monitoring hand. After that, Deep Neural Networks (DNNs) have been applied by them for automatic feature studies that were not sensitive to motion, scaling, and rotation from hand-held posture images.[14,15] HGR system focused on application and vocabulary independent for cognitive mapping of different gestures to the same command. The main objective of this system is to recognize hand gesture to control multiple applications.[16]

As shown in the examples above, HGR systems generally use mixed and complex methods. This paper describes how to do hand gesture recognition easily and quickly with a well-tuned deep CNN, without even entering the hybrid techniques. The rest of the paper continues as follows. The methods used are described in Sec. 2 and the features of the data set are given in detail. In Sec. 3, the application of HGR is explained. The results obtained in Sec. 4 are presented.

1.1. *Motivation of the proposed work*

The motivation behind the research is to make an interaction between human and machine. It recognizes emotion by just using some basic shapes in hand to express some ideas which are hard to verbalize. Our aim is to develop a system which will be able to detect gestures that are widely applied in various applications of human–computer interaction. Despite of a number of solutions explained in this literature, the problem with hand tracking remains still far from being solved as there is an important amount of articulation and self-occlusion which cause difficulties in the existing algorithms.It is seen that the proposed CNN model is a suitable method for hand gesture recognition. It achieves high performance measures at less computational expensive. The experimental result proves that the system proposed is well suited for subjects of any age group and as it can be accessed at any time of the day in any circumstances, so it would be very beneficial to all. The comparison with the existing works in the domain proves that our proposed work is flexible and best suited in its concerned domain.

2. Material and Methods

In this study, discussed about the CNN is used in hand gesture recognition system. In this section, the working principle of static hand gesture recognition system has been described. When the input hand data set deals, the feature extraction from the region of interest is determined. In this work, there is no separate image preprocessing and feature extraction is required. It is carried out direct in CNN with good response of the feature values of hand gesture data.

2.1. *Convolutional neural network*

CNN is an alternative neural network type that can be used to reduce spectral variations and spectral correlation models in sound signals. CNNs are structurally similar to feed-forward neural networks. CNNs establish a close bond between the nodes in the adjacent layers, allowing them to cross attribute representations. That is, the entries of the hidden nodes on a layer are the data sets formed by the corresponding nodes in the previous layer. Therefore, combining so many layers results in nonlinear spiral filtering. In a CNN, the convolutional layer filters the input images and filters. These nodes generate maps using shared weights and prejudices. Gradient methods can be used to learn shared weights and biases.[17,18]

CNN architectures can be of many different structures. It consists mostly of sublayers that transmit and regulate properties together with the convolutional layer. As in a standard feed-forward neural network, one or more fully bound layers that are required for classification or regression complement these layers. Naturally, the layers are sequentially stacked to create a deep pattern. Figure 1 shows the typical CNN model applied to classify. An image that enters the CNN, first enters the convolutional layer, then is sampled in the combining layer called pooling. The samples from these processes then feed one and sometimes more than one fully bound

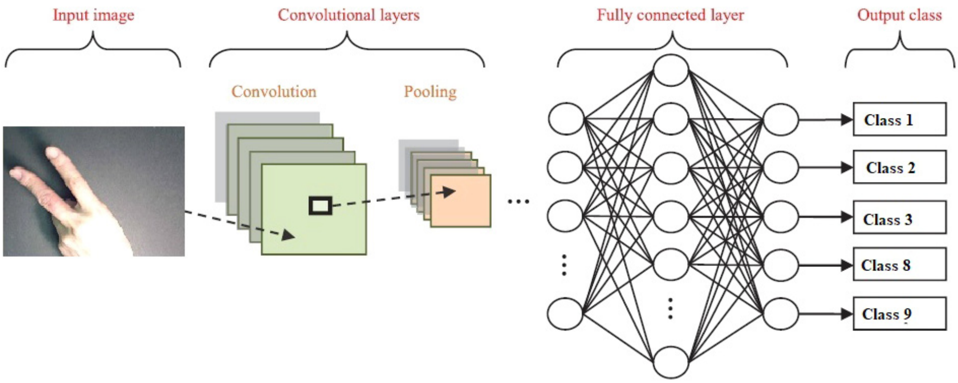

Fig. 1. The typical CNN architecture applied to classify image.[9]

layers. In conclusion, the classification layer subtracts the label of the class. Although this is the most common basic design in the works, various architectural modifications have been planned in recent ages to improve image classification accuracy or to reduce calculation costs.[9]

2.2. *Cambridge hand gesture data set*

The data set of experiments is created on the Cambridge Hand Gesture data set.[19] The data set involves a series of 900 images consisting of 9 movement classes describing three original hand figures and three original movements. So, the target task of this data set is to classify different shapes and different movements parallelly.[20] There is a total of 100 images for each class, with a total of five different illumination, 10 arbitrary moves, and 900 images for two subjects. Every sequence was noted in front of a stationary camera with roughly remote movements in space and time. Therefore, large intra-class changes in spatial and temporal alignment are observed in the data set. Typical example sequences of 9 classes for five different lighting prototypes are given in Fig. 2.[20]

2.3. *Performance evaluation*

The accuracy (Ac), sensitivity (Se), and specificity (Sp) are commonly used performance measures in medical classification studies. These measures were used to find out the precision of the proposed model. They are calculated as the following equation[17,21]:

$$\mathrm{Ac} = \frac{T_P}{T_P + F_P}, \tag{1}$$

$$\mathrm{Se} = \frac{T_P}{T_P + F_N}, \tag{2}$$

$$\mathrm{Sp} = \frac{T_N}{F_{PTN} + T_N}, \tag{3}$$

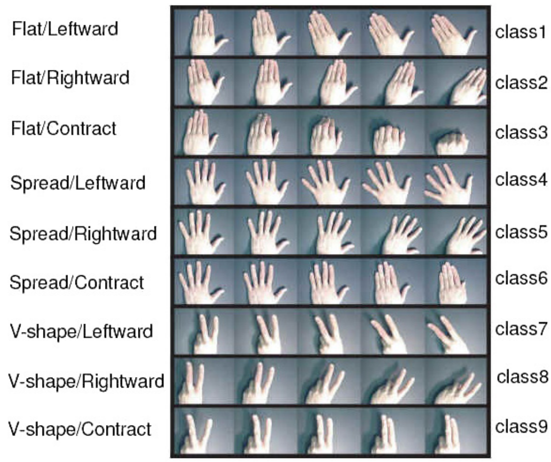

Fig. 2. Sample sequences of the 9 gesture classes.

where T_P and F_P represents the measure of true positives and false positives, and T_N and F_N are the measures of true negatives, and false negatives, respectively. F_{PTN} also represents the false positives and is calculated from negative samples in the results of classification.

The precision of the classifier's ability to diagnose correctly is determined by the ratio of accuracy. The range to which the model properly describes the formation of the target class is defined by the rate of Sensitivity. The extent of the model's target class separation capability is defined by the rate of Specificity.[22,23]

3. Application of Hand Gesture Recognition

The Cambridge Hand Gesture data set was used for testing the performance of the convolutional neural network. The images of the data set, 80% were taken as test data and 20% were taken as test data. The default maximum age in CNN is 1000. The structural details of the CNN used for classification are given in Fig. 3. A total of eight layers were used in CNN and the features are described in the following details.

The input layer to CNN indicates the values of the image inputs. The $240 \times 320 \times 3$ value here indicates that the image height is 240 pixels, the width is 320 pixels, and the third is in the RGB color space. Any data transformation, like data normalization or data increment on this layer, can be considered as random dialing and mapping of data. They are usually used to prevent overfilling and are automatically performed at the initial stage of training.

In CNN, one convolutional layer has been applied. The convolutional layers extract the properties of the image inputs. Thus, they learn the representative properties of image inputs. Each node in convolutional layers is arranged as maps of these representative features. These nodes are tightly connected to each other and can also

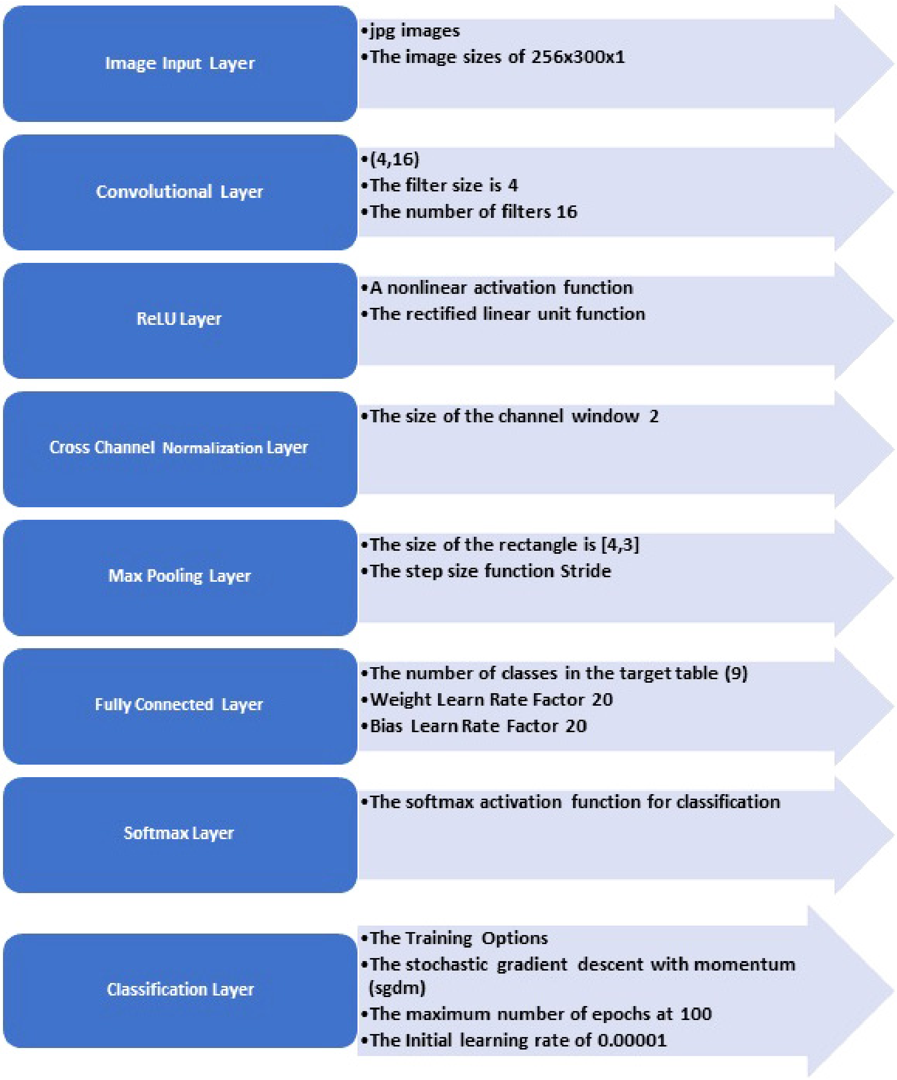

Fig. 3. The structural details of the CNN.

serve as a filter bank by connecting the training weights of the nodes in the previous layer. The parameters of the convolutional layer given herein are in fact the number of neurons that determine the size of the network filter and the number of feature mappings and which are connected to the same area of the output. Four network input images are the filter height and width when scanning. The number of neuronswas 16.

One reLU layer was used in CNN. A rectified linear unit function was applied in the CNN. In this embodiment across channel normalization layer was used.

The channel window size is kept as 2. In other words, the size of the channel window to normalize. Pool layers are intended to decrease the spatial resolution of the property maps and therefore provide spatial variability in input distortion and translation. Maximum pooling layers spread the greatest value in the receiving area to the next layer.[24] The maximum pooling layer is used as a further way to reduce the number of parameters for down sampling and to prevent over-sitting. This layer returns the greatest values of the rectangular areas of the entries mentioned by the pool size of the first argument. In this application, the size of the rectangle is (4, 3). The step function is also used to specify the step function when the exercise function is scanning the image. The fully connected layer consists of every feature that the preceding layers have learnt throughout the image to identify greater patterns. The final fully connected layer merges them to classify. Therefore, the size of the output in the latest completely connected layer is equivalent to the number of classes in the target table. In this application, the output size is as mentioned above 9. The soft-max activation function has been applied to classify the fully connected layer. Classification layer is the last one of CNN. This layer applies the possibilities of the said activation function to allocate one of the special classes for each of them.[17]

4. Results and Discussion

The experimental results obtained with the proposed work are shown in Table 1, consist of accuracy, recall, precision, and F1 score.[25] It shows better results of average success rate of 97%. For CNN1 has only two layers of convolution, accuracy of 96%, and for CNN 2, 3, and 4 accuracy maintained above 97%.

It emphasizes that initial point from three layers of convolution, incorporate with pooling layers; there is significant change in the neural network architecture. It does not increase the feature extraction and classification process of the model. So therefore, there is no much increase in accuracy, but computational complex is too high. The deep network architecture listed in Table 2 is the performance measures such as accuracy, recall, precision and F1 score. However, the existing network architecture is too complex than the proposed work, since it has more number of deep layers in the network model.

The classification of the Cambridge Hand Gesture data set has to make with CNN model. It is possible to extract the features and classify gesture patterns to reach high accuracy and robustness. Tables 3 and 4 show the different layers and network architecture performance measures.

Table 1. Performance measures of each layer in CNN.

Measures	CNN 1 (%)	CNN 2 (%)	CNN 3 (%)	CNN 4 (%)
Accuracy	96.00	97.05	97.23	97.66
Recall	96.51	97.21	97.65	97.86
Precision	96.45	97.12	97.87	97.45
F1 score	96.35	97.43	97.89	97.25

Table 2. Comparison of the performance measures with different network architectures.

Measures	Inception ResNetV2 (%)	ResNet 50 (%)	Dense Net201 (%)	LeNet (%)
Accuracy	98.95	96.99	97.54	92.54
Recall	98.56	96.56	97.77	92.61
Precision	98.76	96.87	97.94	92.53
F1 Score	98.54	96.77	97.63	92.76

Table 3. Performance measures of each layer in CNN.

	CNN 1 (%)	CNN 2 (%)	CNN 3 (%)	CNN 4 (%)
Cambridge Hand gesture data	96.73	96.47	97.23	97.56
Barczak *et al.*[26]	97.65	97.89	97.55	98.96
Moeslund's[27]	98.54	98.66	98.78	98.65

Table 4. Performance measures with different network architectures with respect to hand gesture data.

	Inception ResNetV2 (%)	ResNet 50 (%)	Dense Net201 (%)	LeNet (%)
Cambridge Hand gesture data	98.45	97.69	98.64	88.56
Barczak *et al.*[26]	98.66	97.86	97.57	98.61
Moeslund's[27]	98.46	97.67	98.54	97.53

The classification experiments were iterated 20 times and the average accuracy, sensitivity and specificity values were calculated. The accuracy was found to be 96.66%, the sensitivity was measured as 85%, and specificity was found to be 98.12% according to the average values obtained at the end of 20 times of operation. The accuracy level of the proposed work is better than other state-of-the-art methods shown in Table 5. However, the sensitivity and specificity values are varied for other

Table 5. The comparison data of the previous handgesture recognition works.

No.	Classification method	Overall accuracy (%)	Sensitivity (%)	Specificity (%)
1	CNN **(Proposed)**	**96.66**	**85.00**	**98.12**
2	GPF[28]	85.00	87.75	89.8
3	HDN[29]	85.60	83.55	88.15
4	AFMKL[30]	87.27	96.4	97.52
5	COV3D[31]	93.91	82.12	85.65
6	HMM + DTW[32,37]	93.98	80.13	90.25
7	DSG + AR[33]	94.00	83.33	94.5
8	CCA[34]	82.00	97.6	97.2
9	TB[35]	93.00	78.8	87.4
10	HOSVD[36]	90.10	81.5	86.5
11	JSR[36]	87.75	77.5	88.33
12	JSR + LLS[36]	93.80	—	—
13	GM[36]	95.10	—	—
14	SANS[36]	96.40	—	—

methods. The classifier model is validated through the accuracy value obtained from the confusion matrix.

In order to evaluate the success of the classification, the obtained results were compared with the results from same database in the past. These are: (1) used CNN in this study, (2) Liu and Shao used Genetic programming generated features (GPF),[28] (3) Kovashka and Grauman used Hierarchy of discriminative space-time neighborhood features (HDN),[29] (4) Wu *et al.* used augmented features in conjunction with multiple kernel learning (AFMKL),[30] (5) Sanin *et al.* used Spatial-temporal covariance descriptors (COV3D),[31] (6) Barros *et al.* used Hidden Markov Models and Dynamic Time Warping (HMM+DTW),[32] (7) Baraldi *et al.* used distributed self-gesture and artwork recognition methods (DSG+AR),[33] (8) Kim and Cipollo used Canonical Correlation Analysis (CCA),[34] (9) Lui and Beveridge used a tangent bundle on a Grassmann manifold (TB),[35] (10) Also Lui and Beveridge used modified High Order Singular Value Decomposition (HOSVD),[36] (11–14) Chen *et al.* used Joint Sparse Representation (JSR), Grassmann Manifolds (GM) and Local Linear Subspace (LLS) and Sparse Approximated Nearest Subspaces (SANS).[36] The methods used in these studies and the obtained results were given in Table 5. The finest results in this table were shown as bold.

As seen from Table 5 and Fig. 4, the CNN method proposed by us has the best accuracy in classification which is 96.60%. Then, the second order is Sparse Approximated Nearest Subspaces (SANS) method with 96.40% accuracy. Unfortunately, specificity, and sensitivity measures were found for a fewer number of previous studies. Considering the existing values, it is seen that the proposed CNN design is the most suitable method for hand gesture recognition. Since the image does not require a method such as pre-processing, optimization or segmentation, the workload of the HGR process is considerably reduced. For example, Islam *et al.*[38] extracted the properties of images with CNN, and then used it to define hand signals with the Multi-Class Support Vector Machine. Using two separate classifiers in image classification means spending extra time and workload. This reveals another advantage of classification with the classic CNN is recommended.

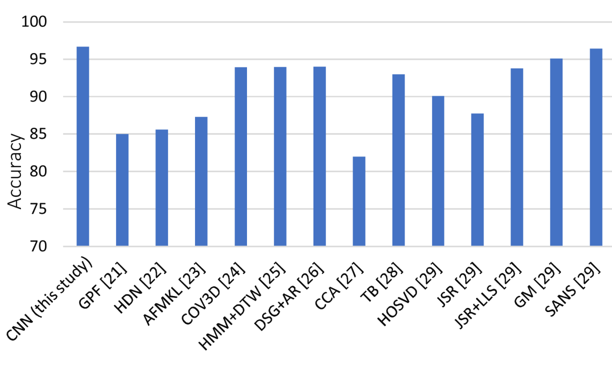

Fig. 4. Comparison of our proposed method with Refs. 28–36.

5. Conclusions

Hand gestures are of great importance in people's communication. For this reason, automatic perception of hand movements within the scope of computerized vision is of great importance. In this study, the classification of CNNs is explained with the aim of recognizing hand gestures in order to cause too much work load with mixed methods. The results obtained from the proposed CNN structure show that the method works very efficiently. In addition, the low workload indicates that the HGR process can be performed more quickly and easily. The main idea here is that it is more important to optimize the settings of the classification method used before classifying methods in the classification studies. However, in subsequent studies, perhaps better classification performance can be achieved by adding other layers (for example, second convolutional layer) to the CNN or by rearranging existing settings. CNN alone has been able to show better performance than the existing hybrid methods in this area. The main contribution of the proposed work, having 96.66% accuracy and comparison with the existing literature proves that it is best suited in its concerned domain. This work can be enhanced in real-time with the dynamic hand gesture using deep neural network approaches.

References

1. H. Francke, J. Ruiz-del-Solar and R. Verschae, "Real-time hand gesture detection and recognition using boosted classifiers and active learning," in *Pacific-Rim Symp. Image and Video Technology* (Springer, Berlin, Heidelberg, 2007), pp. 533–547.
2. E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," *Eng. Appl. Artif. Intell.* **22**(8), 1141–1158 (2009).
3. P. Barros, N. T. Maciel-Junior, B. J. Fernandes, B. L. Bezerra and S. M. Fernandes, "A dynamic gesture recognition and prediction system using the convexity approach," *Comput. Vis. Image Underst.* **155**, 139–149 (2017).
4. P. K. Pisharady and M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," *Comput. Vision Image Underst.* **141**, 152–165 (2015).
5. K. Martin Sagayam and D. Jude Hemanth, "Hand posture and gesture recognition techniques for virtual reality applications: A survey," *Virtual Reality* **21**(2), 91–107 (2017).
6. C. Wang, Z. Liu, M. Zhu, J. Zhao and S. C. Chan, "A hand gesture recognition system based on canonical superpixel-graph," *Signal Process.: Image Commun.* **58**, 87–98 (2017).
7. A. Hariaa, A. Subramaniana, N. Asokkumara, S. Poddara and J. S. Nayak, Hand gesture recognition for human computer interaction, *Procedia Comput. Sci.* **115**, 367–374 (2017).
8. N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Trans. Instrum. Meas.* **60**(11), 3592–3607 (2011).
9. W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.* **29**(9), 2352–2449 (2017).
10. S. Shin and W. Sung, "Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks," *2016 IEEE Int. Symp. Circuits and Systems (ISCAS)* (IEEE, 2016), pp. 2274–2277.

11. A. Dadashzadeh, A. T. Targhi and M. Tahmasbi, "HGR-net: A two-stage convolutional neural network for hand gesture segmentation and recognition," arXiv:1806.05653.
12. P. Molchanov, S. Gupta, K. Kim and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops* (IEEE, 2015), pp. 1–7.
13. V. John, A. Boyali, S. Mita, M. Imanishi and N. Sanma, "Deep learning-based fast hand gesture recognition using representative frames," in *2016 Int. Conf. Digital Image Computing: Techniques and Applications (DICTA)* (IEEE, 2016), pp. 1–8.
14. A. Tang, K. Lu, Y. Wang, J. Huang and H. Li, "A real-time hand posture recognition system using deep neural networks," *ACM Trans. Intell Syst. and Technol. (TIST)*, **6**(2), 21 (2015).
15. A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," *IEEE Access* **6**, 13949–13957 (13957).
16. S. Swarup Rautaray and A. Agrawal, "Hand gesture recognition towards vocabulary and application independency," *Int. J. Image Graphics* **13**(2), 1340001 (2013), https://doi.org/10.1142/S0219467813400019.
17. O. Deperlioglu, "Classification of phonocardiograms with convolutional neural networks," *Broad Res. Artif. Intell. Neurosci.:* **9**(2), 22–33 (2018).
18. O. Deperlioglu, "Classification of segmented phonocardiograms by convolutional neural networks," *Broad Res. Artif. Intell. Neurosci.* **10**(1), 5–12 (2019).
19. T. K. Kim, S. F. Wong and R. Cipolla, "Tensor canonical correlation analysis for action classification," in *CVPR'07 IEEE Conf. Computer Vision and Pattern Recognition*, IEEE, 2007. pp. 1–8.
20. Cambridge Hand Gesture data set, https://labicvl.github.io/ges_db.htm, last accessed on 15.01.2019.
21. A. Gharehbaghia, M. Borgaa, B. J. Sjöbergc and P. Ask, "A novel method for discrimination between innocent and pathological heart murmurs," *Med. Eng. Phys.* **37**, 674–682 (2017).
22. H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Expert Syst. Appl.* **35**, 82–89 (2008).
23. K. Martin Sagayam and D. Jude Hemanth, "ABC algorithm-based optimization of 1-D hidden Markov model for hand gesture recognition applications," *Comput. Industry* **99**, 313–323 (2018).
24. W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural comput.* **29**(9), 2352–2449 (2017).
25. T. Aora and R. Dhir, "Geometric feature based classification of segmented human chromosomes," *Int. J. Image Graphics*, **20**(1), 2050006, (2020), https://doi.org/10.1142/S0219467820500060.
26. A. L. C. Barczak, N. H. Reyes, M. Abastillas, A. Piccio and T. Susnjak, "A new 2D static hand gesture colour image dataset for ASL gestures," *Res.Lett. Inf. Math. Sci.* **15**(4356), 12–20 (2011).
27. T. B. Moeslund and E. Granum, Modelling and estimating the pose of a human arm, *Mach. Vis. Appl.* **14**(4), 237–247 (2003).
28. L. Liu and L. Shao, "Synthesis of spatio-temporal descriptors for dynamic hand gesture recognition using genetic programming," in *10th IEEE Int. Conf. and Workshops Automatic Face and Gesture Recognition (FG)*, IEEE, 2013, pp. 1–7.
29. A. Kovashka and K. Grauman, "Learning a hierarchy of discriminative space-time neighborhood features for human action recognition," in *2010 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, (IEEE, 2010), pp. 2046–2053.

30. X. Wu, D. Xu, L. Duan and J. Luo, "Action recognition using context and appearance distribution features," in *Proc. 2011 IEEE Conf. Computer Vision and Pattern Recognition* (IEEE Computer Society, 2011), pp. 489–496.
31. A. Sanin, C. Sanderson, M. T. Harandi and B. C. Lovell, "Spatio-temporal covariance descriptors for action and gesture recognition," in *2013 IEEE Workshop on Applications of Computer V'ision (WACV)*, IEEE, 2013, pp. 103–110.
32. P. Barros, N. T. Maciel-Junior, B. J. Fernandes, B. L. Bezerra and S. M. Fernandes, "A dynamic gesture recognition and prediction system using the convexity approach," *Computer Vision Image Underst.* **155**, 139–149 (2017).
33. L. Baraldi, F. Paci, G. Serra, L. Benin and R. Cucchiara, "Gesture recognition using wearable vision sensors to enhance visitors' museum experiences," *IEEE Sens. J.* **15**(5), 2705–2714 (2015).
34. T. K. Kim, S. F. Wong and R. Cipolla, "Tensor canonical correlation analysis for action classification," in *2007 IEEE Conf. Computer Vision and Pattern Recognition*, IEEE, 2007, pp. 1–8.
35. Y. M. Lui and J. R. Beveridge, "Tangent bundle for human action recognition," in *2011 IEEE Int. Conf. Automatic Face & Gesture Recognition and Workshops (FG 2011)*, IEEE, 2011, pp. 97–102.
36. S. Chen, C. Sanderson, M. T. Harandi and B. C. Lovell, "Improved image set classification via joint sparse approximated nearest subspaces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (IEEE, 2013), pp. 452–459.
37. K. Martin Sagayam and D. Jude Hemanth, "A probabilistic model for state sequence analysis in hidden Markov model for hand gesture recognition," *Comput.Intell.* **35**(1), 51–81 (2019).
38. M. R. Islam, U. K. Mitu, R. A. Bhuiyan and J. Shin, "Hand gesture feature extraction using deep convolutional neural network for recognizing American sign language," *2018 4th Int. Conf. Frontiers of Signal Processing (ICFSP)*, Poitiers, 2018, pp. 115–119, doi: 10.1109/ICFSP.2018.8552044.

K. Martin Sagayam, received his PhD in Electronics and Communication Engineering (Signal image processing using machine learning algorithms) from Karunya University. He received Master of Engineering in Communication Systems from Anna University, Chennai and received Bachelor of Enigeering in Electronics and Communication Engineering from Anna University. Currently, he is working as Assistant Professor in the Department of ECE, Karunya Institute Technology and Sciences, Coimbatore, India. He has authored/ co-authored more number of referred International Journals. He has also presented more than 20 papers in reputed international and national conferences. He has authored 2 edited book and 10 book chapters with reputed international publishers like Elsevier, Springer, IGI Global and CRC press. He has editorial board in the reputed book and book series with reputed international publishers like IGI Global, and CRC Press. He has reviewed research articles from the journal of Signal, Image and Video Processing, Intelligent Decision Technologies, International Journal of Engineering Research & Technology, Pattern analysis and its applications and Informing Science Institute. He is an active member of professional bodies such as Engineering and Scientific Research Groups, International Society of Promising Computer Engineers, Copernicus, Scientific Engineering Research Corporation, International Association of Computer Science and Information Technology, International Association of Engineers, Indian Society of Electronics and Communication Engineering and Orcid. His area of interest includes Communication systems, signal and image processing, machine learning and virtual reality.

Diana Andrushia received her Bachelor of Engineering degree in Electronics and Communication Engineering with first class from Anna University, Chennai, India in 2006 and her Master of Engineering degree in Applied Electronics with first class from Anna University, Chennai, India in 2008 and her Doctoral degree in Information and Communication Engineering from Anna University, Chennai, India in 2018. Currently, she is working as Assistant Professor, in the department of ECE, Karunya University Coimbatore. She has authored 35 research articles in reputed journals and conferences. She has reviewed the research articles from reputed journals like, IEEE Access and IEEE Transactions on Neural Networks and Learning Systems. Her research interest includes, attention modeling, artificial intelligence, optimization techniques and deep neural networks.

Ahona Ghosh is a B.Tech, M.Tech in CSE and presently pursuing her research in the area of Machine Learning and Internet of Things from Maulana Abul Kalam Azad University of Technology, West Bengal (In house). She has published more than 20 papers in International conferences, Peer reviewed international journals and book chapters with IEEE, Springer, Elsevier, IGI Global, CRC Press etc. Her research interests include Cloud computing and Blockchain also.

Omer Deperlioglu received his BSc in Electric and Electronic (1988) Gazi University, MSc in Computer Science (1996) from Afyon Kocatepe University, PhD in Computer Science (2001) from Gazi University in Turkey. Now he is associate professor of Computer Programming in Department of Science, Vocational School of Afyon, and Afyon Kocatepe University of Afyonkarahisar, Turkey. His current research interests include different aspects of Artificial Intelligence applied in Power Electronics, Biomedical, and Signal Processing. He has edited 1 book and (co-) authored 3 books and more than 20 papers, more than 30 conferences participation, member in International Technical Committee of 4 conferences and workshops.

Ahmed A. Elngar he ‘received his Ph.D from Faculty of Computers and Artificial Intelligence. Computer Science Department Beni-Suef University. Dr. Elngar is the Founder and Head of Scientific Innovation Research Group (SIRG) and Assistant Professor of Computer Science at the Faculty of Computers and Artificial Intelligence, Beni-Suef University. Dr. Elngar is a Director of the Technological and Informatics Studies Center (TISC), Faculty of Computers and Information, Beni-Suef University. Dr. Elngar has more than 25 scientific research papers published in prestigious international journals and over 5 books covering such diverse topics as data mining, intelligent systems, social networks and smart environment. Dr. Elngar is a collaborative researcher He is a member in Egyptian Mathematical Society (EMS) and International Rough Set Society (IRSS). His other research areas include Internet of Things (IoT), Network Security, Intrusion Detection, Machine Learning, Data Mining, Artificial Intelligence. Big Data, Authentication, Cryptology, Healthcare Systems, Automation Systems. He is an Editor and Reviewer of many international journal around the world.

Dr. Elngar won several awards including the “Young Researcher in Computer Science Engineering”, from Global Outreach Education Summit and Awards 2019, on 31 January 2019 (Thursday) at Delhi, India. Also, he awards “Best Young Researcher Award (Male) (Below 40 years)”, Global Education and Corporate Leadership Awards (GECL-2018), Plot No-8, Shivaji Park, Alwar 301001, Rajasthan.