International Journal of
**INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING**

ISSN:2147-6799 www.ijisae.org Original Research Paper

# An Intelligent System for Recognizing the Human Activity using Improved Convolutional Neural Network (I-CNN) Model

**A. Venugopal Rao [1], Santosh Kumar Vishwakarma [2], Shakti Kundu [3]***

**Submitted**: 24/09/2023 **Revised**: 17/11/2023 **Accepted**: 27/11/2023

**Abstract** In account of the many practical uses of computer vision, researchers have made human behavior detection in videos a top priority. It covers a broad range of topics, such as video surveillance, behavior analysis, sports analysis, e-health, patient monitoring, assisted daily living, and much more. Many researchers have been proposed methods that rely on vision to identify human activity. Researchers will need to address problems of variations in human activity detection, class comparison between pictures, and temporal variation in order to construct an effective vision-based human behavior detection system.

This paper proposes an improved Convolutional Neural Network (I-CNN) to boost human behavior detection performance and address these issues. A dilated convolution filter widens the convolutional layer's receptive field to allow for the incorporation of more pertinent data. Additionally to the standard CNN model, it boosts CNN performance and reduces the calculation time. We've compiled KTH publicly available benchmark dataset, to test the effectiveness of suggested process. We show experimental findings showing that our suggested model outperforms the standard CNN model on the aforementioned dataset, achieving 96.85%.

***Keywords:*** *Human activity, Improved CNN, deep learning, activity recognition and artificial intelligence.*

## 1. Introduction

In the research area of computer vision, human behavior detection in videos have become more important as a research goal [1]. It covers a wide range of fields, from medicine to security to sports analytics to patient monitoring. Recognizing activities is the process of labelling the actions shown in a video (a series of still images) as belonging to one of several categories [2]. Indoor and outdoor pursuits, daily routines, outings to restaurants and stores, and so on are just a few examples of the many types of activities that a person engages in. Figure 1 shows two main types of Identification of human activities: those based on vision and those based on sensors [3].

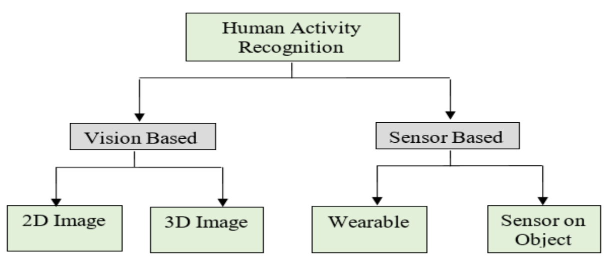


**Fig. 1** Types of human behavior detection

The development of an autonomous system that is capable of effectively recognising and comprehending human activities and behaviours is one of the most important goals of the artificial intelligence civilization. To enhance its potential to benefit society, a robot assistant may be engineered to perform tasks like providing support to a patient while they are being observed at home, determining the appropriate workout regimen, and shielding the patient from additional harm. [28][4]. As a result, such a sophisticated system will be of great assistance to us because it will shorten the amount of time spent in the doctor's office, lower the cost of medical care, and enable constant remote monitoring of the patient.

Over the course of the previous two decades, numerous feature-based techniques, both manually produced and automatically taught, have been evolved for the purpose of Recognition of Human Action in videos. The first attempts at recognising human activities were made using hand-crafted features that concentrated mostly on fundamental, elementary behaviours [5,6]. These are the four pillars of the approach. It has been found that handcrafted feature solutions produced promising results, although these solutions depended more about action classification using feature descriptors. These solutions necessitated additional work and subject matter expertise. [7-9].

**Organization of the paper:** This research paper is comprised of the following five major sections: Section 2, presents an extensive literature review of vision-based human behavior detection , identifies research gaps, and then discusses the purpose of the research. The improved convolutional neural network model that has been proposed will be explained in Section 3. In section 4, the

[1] *Research Scholar, Manipal University Jaipur, India.*
*vengopal.229351007@muj.manipal.edu*
[2] *School of Computer Science & Engineering, Manipal University Jaipur. India.*
*santosh.kumar@jaipur.manipal.edu*
[3] *Directorate of Online Education, Manipal University Jaipur , India.*
*shakti.kundu@jaipur.manipal.edu*

results of the experiments relevant to the proposed improved CNN model are discussed and compared with traditional CNN model. In the Section 5, we will discuss the findings and give a summary, as well as recommended directions for further research.

## 2. Literature Survey

Many manually constructed and automatically learnt feature-based methods for recognising human actions in video sequences have been created in the previous few decades. Feature extraction, feature representation, and feature classification are the three steps which are involved in human baviour detection. Using a hand-crafted feature extractor, low-level features are extracted in the first step. To make the information more distinctive for the final feature classification, certain feature representation algorithms are then used to encode it [10][29].

The core concept behind feature extraction can be broken down into two distinct groups: those that are based on worldwide features and those that are based on local features. In the first approach, the representation is treated as a whole, and background subtraction or tracking must be used to identify the area of interest. In order to make the observation stand out more for the final feature classification, local representation is presented in the form of patches that are sampled either densely or at space-time interest spots [11][30-32]. It more unique feature models can improve practical classification results.

It was believed that the distinguishability of the feature representation was the primary factor in determining the level of performance achieved by the hand-crafted recognition techniques. As a result, one of the essential components of conventional techniques for action recognition is rewriting the initial feature using discriminative words. In recent times, numerous feature representation algorithms have been suggested, and many methods, such as BOW [12] and VLAD [13] [33], demonstrate effective performance in feature encoding. Unfortunately, because of the limitations imposed There are still certain challenges in hand-crafted feature processing that have not been overcome by human cognition that are relevant to the final feature classification.

To begin, each widely employing a feature extractor is developed according to particular databases; as a result, The extractors of features are frequently biased towards the databases they were intended for and cannot have the capability to extract features for general purposes [14]. Second, Careful feature engineering was needed to build a handcrafted human behavior system. A deep learning based approach, in contrast to the traditional three-step architecture, are used [15]. The database-bias issue can be mitigated to some degree by the fact that feature learning is less susceptible to the influence of any one learner.

The automatic learning process based on joint parameters also greatly aids in the exploration of complex structures. As of late, CNNs have proven to rank among the most efficient feature acquisition models for picture-based feature extraction [16]. Normal CNN designs, however, can't handle video data because a 2D filter can't capture 3D features which enhanced the traditional CNN to 3D-CNN, changing the 2D to 3D filter, and hence bringing 2D ConvNet to the video realm [17]. The experimental findings demonstrated that this design can achieve higher performance than the 2D CNN and the custom-made video analysis feature in most situations. To get around the problem of recording motion information between frames, two-stream CNNs are also widely used. Many better models have been suggested by K. Simonyan [18].

Using the UCF101 as training data, Wang suggested a two-stream ConvNets approach that incorporated spatial and motion features and achieved perfect results [19], demonstrating once and for all the significance of temporal data.

**Table 1** Literature Survey of various Human behavior detection

| Author | Methodology | Key Findings | Limitation |
|---|---|---|---|
| B. Saghafi and D. Rajan | Machine Learning Techniques | This proposed model used SIFT feature extraction technique and SVM classifiers. | The model saturate in terms of accuracy. |
| K. Simonyan | The proposed model used two-stream convolutional network approaches. | The proposed model used spatial and motion information features for human activities. | Training time is lead to more. |

**Identification of the gaps:** In order to predict the human activity, following observation have been made:

- ❖ It has been observed that descriptors based on classic machine learning approaches are too inflexible to capture all potential alterations in video frames, such as changes in scale, views, occlusion, etc..
- ❖ Furthermore, the small features inside the silhouettes cannot be found by typical feature extraction based

techniques..

- ❖ It is evident that sudden changes in scenes occur frequently in videos.
- ❖ Some videos have minimal frame-to-frame volatility, while others move excessively quickly. For this reason, extracting key poses frames utilizing standard algorithm because of the risk of low accuracy prediction.

All of the models that were discussed above are inefficient because of delay in processing. Hence CNN model for the human behavior detection after taking into consideration the drawbacks that were discussed earlier.

**Research Questions:**

Here are the challenging research questions addressed in this paper are as follows:

1. How to effectively apply Improved CNN for human behavior detection .
2. How to reduce the training time for human behavior detection dataset.
3. How to increase the prediction accuracy in improved Convolutional Neural Network Model.

**Research Methodologies:** To finish the current study with an existing HAR system, this research activity mostly concentrated on secondary research using existing datasets indicated in state-of-the-art techniques. There is no distinct dataset created to finish this study work. Furthermore, existing earlier literature on HAR has been evaluated in order to comprehend the background work in the associated topic.

This study demonstrates a solution employing deep learning technique approaches with KTH dataset and evaluates it using statistical methods such as accuracy, precision, and recall, confusion matrix, and so on. As a result, the current study can be categorised as empirical research.

**Objective of the Proposed Work:** The main objectives of this research is to develop an develop an automated intelligent system for human behavior detection from static actions (2D images). Also the model reducing the computational complexity while simultaneously increasing the accuracy of human behavior detection .

## 3. Proposed Work

In order to reduce the computational difficulty and boost the precision of human behavior detection, we have presented an Improved CNN that is Dilated CNN. Figure 1 depicts the architecture of the CNN, which serves as the foundation for the Improved CNN. The purpose of an expanded convolution kernel is to expand or widen the receptive field of the kernel without increasing the number of parameters or the gap weight values used in the filters. As a consequence of this, we are able to improve the model's overall effectiveness.

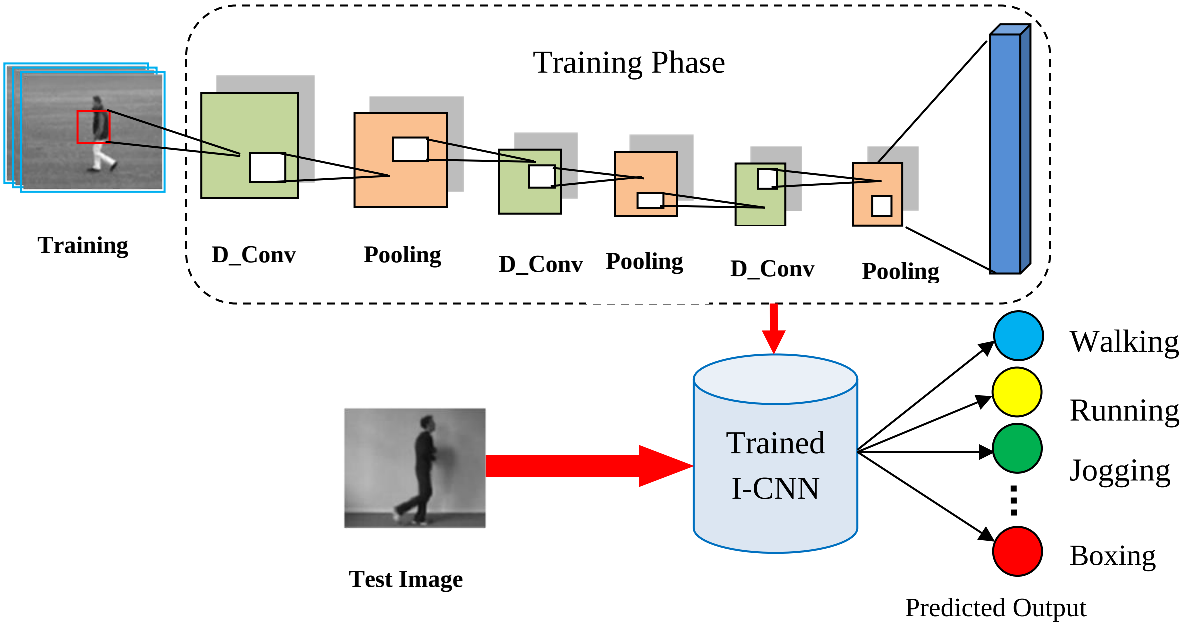


**Fig.** 1 Architecture of Improved-Convolutional Neural Networks

### 3.1 Improved Convolutional Layer

A specific kind of convolutional layer known as the dilated convolutional layer was at first developed with the goal of image segmentation in mind. The use of dilated convolutions in CNN was pioneered by Yu and Koltun [20], who were the first to do so in order to acquire more contexts in image recognition tasks. During the process of convolution, a modified kernel that has a specified dilation rate (dr) of 2, 3 is applied to the image that has been provided as input.

The network's receptive field grows bigger as a result of the dilation rate, which enables the collection of additional information regarding the context. Figure 2 depicts a traditional convolution kernel as well as a dilated convolution kernel over an image with dimensions of 10 by 10. The traditional convolution kernel has a measure of three by three, and the dilated convolution kernels have dimensions of two and three, respectively. However, the total number of parameters in each of the dilated convolution kernels remains the same even though the dilation rate causes an increase in the area of the receptive field. Consequently, utilising a dilated convolution kernel to analyse images will result in the extraction of additional information from the convolution kernel without an accompanying increase in the amount of computation required.

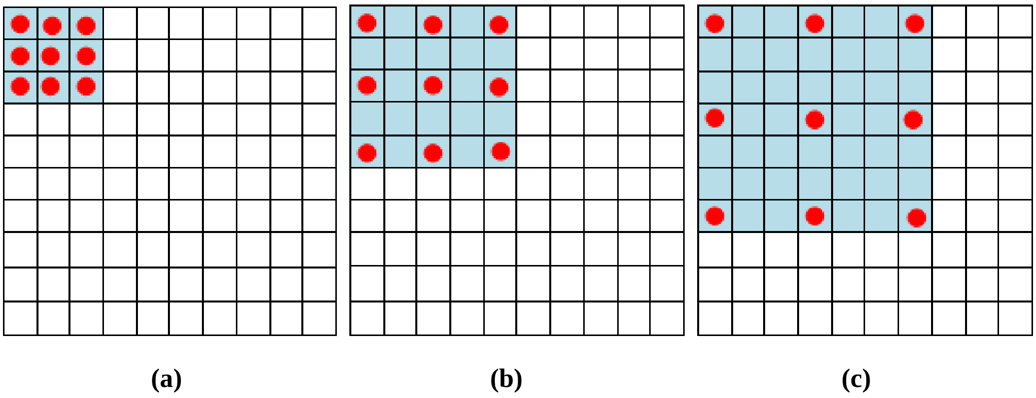


**Fig. 2** Kernel receptive field for traditional & dilated convolutional layer

The receptive field size is generally defined as in equation 1:

$$[k + (k - 1)(dr - 1)] \times [k + (k - 1)(dr - 1)] \quad (1)$$

where k represents kernel size, dr represents the dilation factor of the kernel. Using a 3x3 kernel and dilation factor 1 (dr=1), we get our regular 3x3 receptive field. When dilation factor is 2 (dr=2), a 5x5 receptive field is obtained for a 3x3 kernel, where as when dilation factor is 3 (dr=3), a 7x7 receptive field is obtained. The general form of convolution is defined as:

$$s(i,j) = \sum_{k=1}^{n_{in}} (X_k \times W_k)(i,j) + b \quad (2)$$

where

$n_{in}$ = input matrices of the tensor,

$X_k$ = $k^{th}$ input matrix,

$W_k$ = the $k^{th}$ sub-convolution kernel matrix of the convolution kernel.

s(i,j) = output values for matrix of corresponding elements to the kernel w and b represents bias value.

The dilated convolution process of a input of size 10×10 with padding size 1, 3×3 kernel with dr 2 and stride 1 is illustrated in Figure 3.

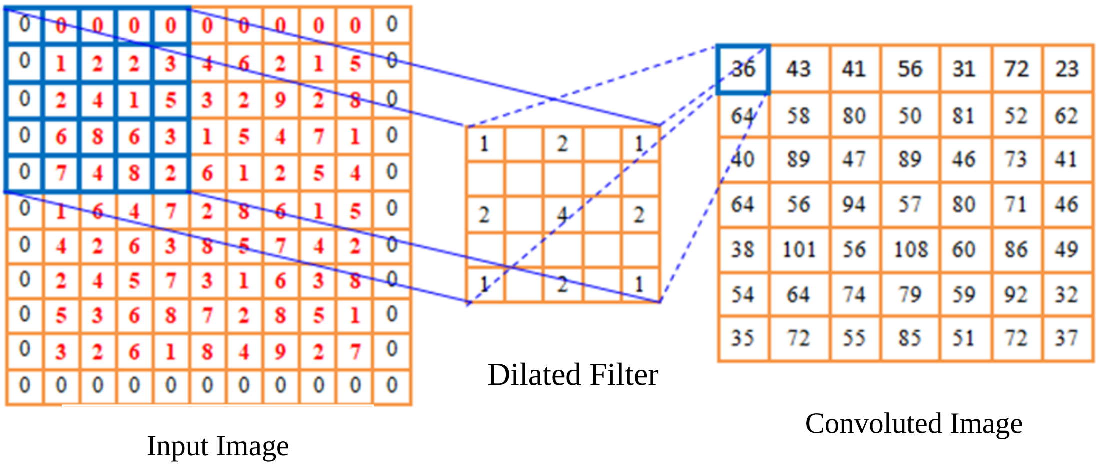


**Fig. 3** Dilated convolutional layer

### 3.2 Pooling Layer

The subsampling process known as the pooling layer is applied after the activation function and convolutional layer. A pooling layer's primary goal is to minimize maps' feature while preserving the crucial data. As an outcome, it lowers the quantity of network computations and parameters that need to be learned. Additionally, it aids in avoiding the overfitting issue. When we select the images as an input matrix I with dimensions I = (Ix, Iy), we can write the output size of the corresponding matrix as O = (x, y) stated in equations 3 and 4.

$$x = \frac{I_x - (W_x - S_x) + 2 \cdot P_x}{S_x} \quad (3)$$

$$y = \frac{I_y - (W_y - S_y) + 2 \cdot P_y}{S_y} \quad (4)$$

where, I = (Ix, Iy) is a two dimensional matrix of the feature map as input image, W = (Wx, Wy) is represented as pooling window size, S = (Sx, Sy) is an stride length which is used to determine by how many pixels the window will be shifted, P = (Px, Py) is padding parameter and O = (x, y) is an output of the pooling result as a two dimensional matrix. There are three common pooling methods, namely max-pooling, min-pooling and average-pooling.

**Max and Min Pooling**

The max-pooling method finds the highest value for every feature map patch. Similarly min-pooling finds out the minimum value for each patch of the feature map. In this case we have the pooling window size W = (2, 2) and the size of stride length S = (2, 2). It means that the size will be reduced to half of the size. If the input matrix for pooling layer is 4×4 matrix, the output of the pooling values will be 2×2 as shown in Figure 4.

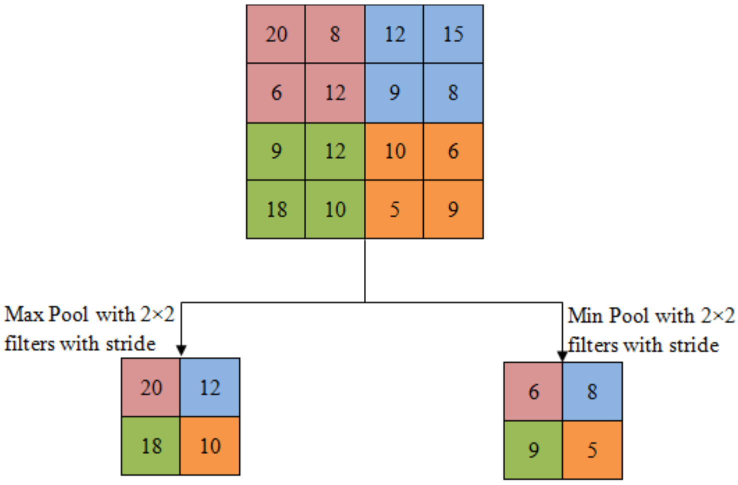


**Fig. 4** Max and Min-pooling operation

### 3.3 Flatten and Fully Connected Layer (FC Layer)

The output of the convolution/pooling technique is used by a fully connected layer to categorize the image into a label. Towards the end of the model, CNN frequently uses one or more completely connected layers. The outcome of all the convolution and pooling layers combined is stacks of feature maps. A completely linked layer is used to flatten a stack of feature maps into a one-dimensional feature space. To summarize those feature maps, all of the flattened neurons are then connected. This layer makes use of its own weighting to obtain the most accurate. Lastly, it figures out the score for each class in view of the neuron's decision on every one of the marks and the champ of that vote is considered as the end. This process is illustrated in Fig. 5, where a 3×3 map is converted into 9 ×1 neurons which are fully connected to 3 neurons.

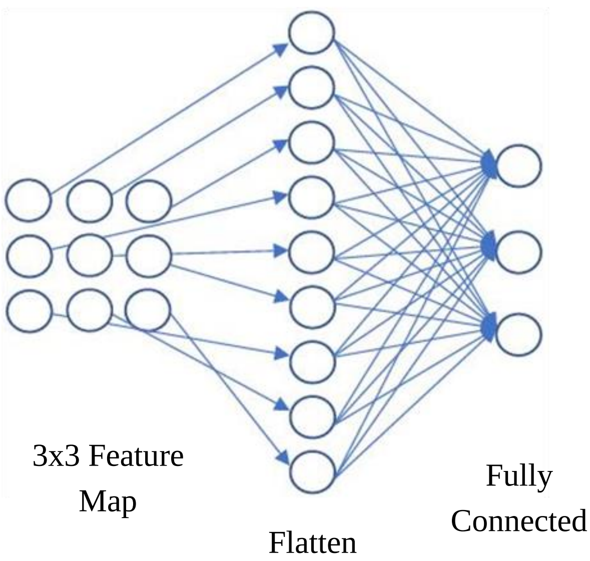


**Fig. 5** Flatten and Fully Connected Layers

## 4. Experimental Results and Analysis

In this section, we have compared the outcomes of proposed experiments designed to demonstrate the usefulness of the traditional CNN model. Experiments are run using deep learning libraries like Open CV [21], Numpy [22], Matplotlib [23], and sklearn [24] in environments like Jupyter Notebook and Anaconda Prompt IDE. Keras [25] and TensorFlow [26] on a corei7 CPU 2.6GHz, 1TB hard disc drive, and 8-GB of RAM were used to train and evaluate the proposed approaches on KTH human activity dataset.

### 4.1 Dataset Collection

In 2004, the Royal Institute of Technology in Sweden was responsible for the creation of the KTH dataset [27]. This dataset is comprised of six distinct human actions, including walking, jogging, running, boxing, hand clapping, and hand waving. These actions were carried out by 25 different actors in four distinct situations. Thus, it includes 25×6×4 = 600 video sequences. These videos were captured using a camera and background that remained still throughout the recording process; as a result, this dataset is considered to be one of the more straightforward options for testing human behavior detection algorithms. Figure 6 depicts a single image representing an example of each action taken in one of four possible situations. As shown in Table 2, 70% of the dataset is employed to train the model, while 20% is used to validate it and 10% is used to test it.

**Table 2** KTH images dataset information

| S. No. | Expression Type | Total No. of Images | Training Images | Validation Images | Testing Images |
|---|---|---|---|---|---|
| 1. | Walking | 1000 | 700 | 200 | 100 |
| 2. | Jogging | 1000 | 700 | 200 | 100 |
| 3. | Running | 1000 | 700 | 200 | 100 |
| 4. | Boxing | 1000 | 700 | 200 | 100 |
| 5. | Hand waving | 1000 | 700 | 200 | 100 |
| 6. | Hand clapping | 1000 | 700 | 200 | 100 |

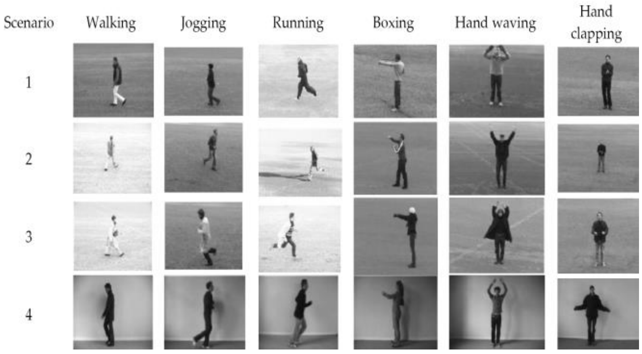


**Fig. 6** One frame example of each action in KTH dataset

### 4.2 Evaluation Metrics

Precision, recall, accuracy, and the F1-measure are only a few of the performance indicators used to assess the suggested model's effectiveness. The aforementioned metrics are computed using the confusion matrix, which is a two-dimensional table as illustrated in Figure 7. Actual values are on the column side and anticipated values are on the row side of this matrix. When the models accurately forecast the positive class, the result is called the TP. When the models accurately forecast the negative class, the result is called TN. When the models predict the positive class erroneously, the outcome known as the FP occurs. It's the FN.

| | P | N |
|---|---|---|
| Y | **True Positive** | **False Positive** |
| N | **False Negative** | **True Negative** |
| | **P** | **N** |

**Fig. 7** Confusion matrix

**Precision**

One of the best metrics for demonstrating how precise a model is is precision. The formula can be used to determine the precision value. Ref Eq (5).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (5)$$

**Recall**

It is employed to determine the proportion of real positives that the model captures and marks as positive. The formula can be used to determine recall value. Refer Eq (6).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (6)$$

**Accuracy**

The number of correctly identified data in a dataset divided by the total number of samples is how one can calculate accuracy, as shown in the equation (7).

$$\text{Accuracy} = \frac{\text{TP+FP}}{\text{TP+FP+TN+FN}} \quad (7)$$

**F1 -measure**

The harmonic mean, or F1-measure, is used to illustrate how the precision and recall metrics are balanced. One can compute the F1-score measure using the equation (8).

$$F = 2 * \frac{\text{Precision} * \text{Recall}}{(\text{Precision}+\text{Recall})} \quad (8)$$

### 4.3 Results and Discussions

In this study, the experimental results of proposed Dilated CNN models are compared with those of a traditional CNN model that has the same parameters and configurations. The configurations and parameters of the dilated convolution model with a dilation rate of 2 are the same as those of the traditional CNN model; however, the receptive field has been increased to 5x5 from 3x3 in the traditional CNN model. The performance of the proposed models improves, while the amount of time required for computation is greatly reduce as a direct result of increasing the area of the receptive field and adding more relevant data. In a similar way, the receptive field is increased as 7x7 for the dilated convolution that has a dilation rate of 3.

Three convolutional layers—max pooling, one flatten, and one completely connect—make up the experimental setup for both the standard CNN and Dilated CNN models. To mitigate the issue of overfitting notions, dropout and Adam optimizers have been employed. The KTH dataset's performance metrics (conventional CNN, dilated CNN-1, and dilated CNN-2 models) are displayed in Figure 8 for a period of 15 epochs on both training and validation data.

**Table 2** Performance Comparison of proposed model with traditional CNN model

| S. No. | Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| 1. | Traditional CNN Model | 92.35 | 92.66 | 92.35 | 92.28 |
| 2. | Dilated rate 2 CNN Model | 94.85 | 95.44 | 94.86 | 94.89 |
| 3. | Dilated rate 3 CNN Model | 96.85 | 97.17 | 96.86 | 96.89 |

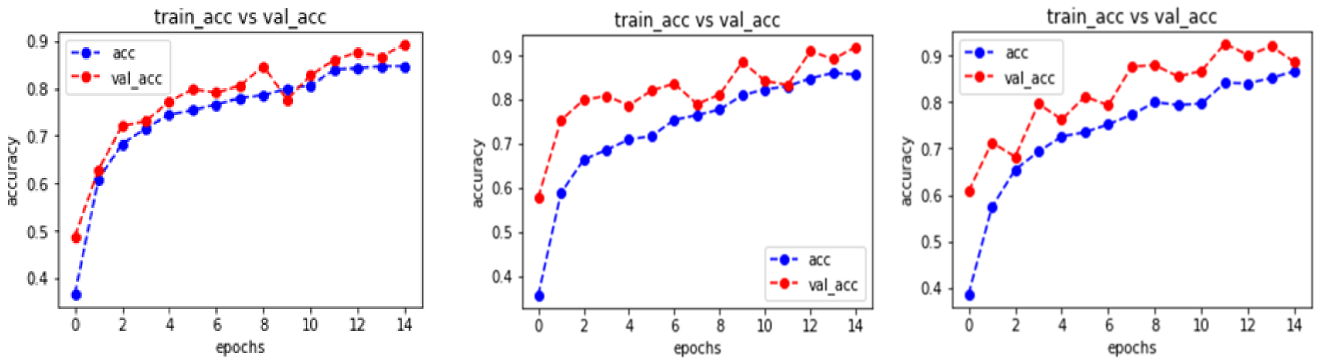


**Fig. 8** Classification accuracy for KTH dataset with Traditional CNN, Dilated 1 and Dilated 2

**Table 3** Confusion matrix of proposed I-CNN model

| | Walking | Jogging | Running | Boxing | Hand waving | Hand clapping |
|---|---|---|---|---|---|---|
| Walking | 95.5 | 0.0 | 2.62 | 0.0 | 0.62 | 0.0 |
| Jogging | 1.0 | 95.25 | 0.0 | 0.0 | 1.25 | 2.5 |
| Running | 0.5 | 0.5 | 95.37 | 1.25 | 2.38 | 0.0 |
| Boxing | 0.0 | 0.0 | 0.0 | 97.5 | 2.5 | 0.0 |
| Hand waving | 0.0 | 0.0 | 0.0 | 1.87 | 94.12 | 4.0 |
| Hand clapping | 0.0 | 1.5 | 1.0 | 0.0 | 0.0 | 97.5 |

In the Dilation rate 2, most of the classes are classified with good results. The following classes such as hand waving, Jogging and Running are classified with poor performance. Most of the hand waving is misclassified as hand clapping, walking is misclassified as running in KTH human activity dataset.

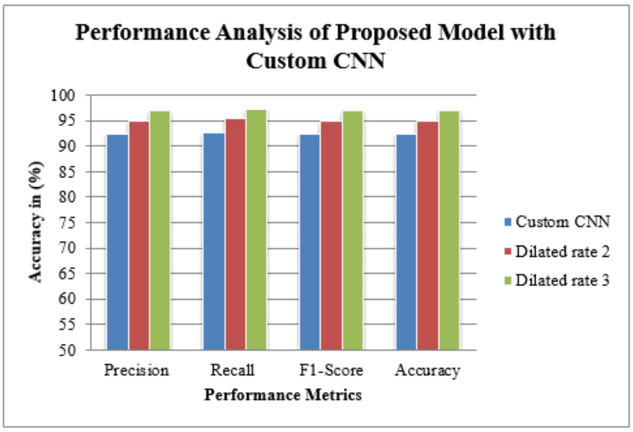


**Fig. 9** Performance analysis of proposed model with

traditional CNN

As shown in Table 3, the performance of proposed model Improved CNN model was compared with existing approaches.

**Table 3** Performance Comparison of proposed model with Existing model

| S. No. | Author | Accuracy |
|---|---|---|
| 1. | B. Saghafi and D. Rajan | 87.35 |
| 2. | K. Simonyan | 91.45% |
| 3. | C. Feichtenhpfer, A. Pinz, and A. Zisserman | 93.25% |
| **4.** | **Proposed Model** | **96.85%** |

## 5. Conclusion

Background information on human action or activity detection in video sequences, significant obstacles to action identification in videos, related HAR applications, and possible implementations utilizing current methods were the starting points for this work. In this paper, Improved Convolutional Neural Network model for human behavior detection of KTH dataset is described. The human activity information was extracted using the dilated convolutional layer and classified using the soft-max classifier. The performance of the proposed dilated convolutional neural network model is evaluated by using KTH benchmark datasets with dilation rate of 2 and dilation rate 3. The accuracy of traditional, dilation 2 and dilation 3 are 92.35%, 94.85% and 96.85% respectively. Experimental results showed better discrimination than traditional CNN in human behavior detection . The experiments were carried out with the help of Python 3.5 and related library packages. In future, we have planned to incorporate our proposed improved convolutional neural network model in human behavior detection system and by implementing in GPU environment for reduce the computational time. Also we have to implement the ensemble CNN model for improve the human behavior detection system in efficient manner.

## References


[1] J. Aggarwal and M. Ryoo, “Human activity analysis: A review,” ACM Comput. Surv., vol. 43, pp. 16:1–16:43, Apr. 2011.

[2] Danafar, S. and Gheissari, N. Action recognition for surveillance applications using optic flow and SVM. In Asian Conference on Computer Vision, Vol. 6(2), pp. 457–466, 2007.

[3] Gorelick, Lena, Moshe Blank, Eli Shechtman, Michal Irani, and Ronen Basri, "Actions as space-time shapes." IEEE transactions on pattern analysis and machine intelligence, Vol. 29, No. 12, pp. 2247-2253, 2007.

[4] Ankur Agarwal and Bill Triggs, “Multilevel image coding with hyper features”, International Journal of Computer Vision, Vol. 78(1), pp. 15–27, 2008.

[5] A. Gilbert, J. Illingworth, and R. Bowden, “Action recognition using mined hierarchical compound features”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33(5), pp. 883–897, 2011.

[6] Dimitris Metaxas and Shaoting Zhang, "A review of motion analysis methods for human Nonverbal Communication Computing", Image and Vision Computing, vol. 31, pp. 421-433, 2013.

[7] Hbali, Y., Hbali, S., Ballihi, L., & Sadgal, M., Skeleton-based human behavior detection for elderly monitoring systems. IET Computer Vision, Vol. 12(1), pp. 16-26, 2018.

[8] Theodorakopoulos, I., Kastaniotis, D., Economou, G., & Fotopoulos, S., Pose-based human action recognition via sparse representation in dissimilarity space. Journal of Visual Communication and Image Representation, Vol. 25(1), pp. 12-23, 2014.

[9] Gaglio, S., Re, G. L., & Morana, M., Human behavior detection process using 3-D posture data. IEEE Transactions on Human-Machine Systems, Vol. 45(5), pp. 586-597, 2014.

[10] S. R. Ke, H. Thuc, and Y. J. Lee, “A review on video-based human behavior detection ,” Computers, vol. 2, no. 2, pp. 88-131, Mar. 2013.

[11] C. Dai, X. Liu, J. Lai, P. Li, and Han-Chieh Chao, “Human behaviour deep recognition architecture for smart city applications in 5G environment” IEEE Netw., DOI:10.1109/MNET. 2019.1800310, 2012.

[12] X. Zhen, and L. Shao, “Action recognition via spatio-temporal local features: A comprehensive study,” Image Vis. Comput., vol. 50, pp. 1-13, Jun. 2016.

[13] H. Kwon, Y. Kim, J. S. Lee, and M. Cho, “First person action recognition via two-stream convnet with long-term fusion pooling,” Pattern Recogn. Lett., vol. 112, pp. 161-167, Sep. 2018.

[14] B. Saghafi and D. Rajan, “Human action recognition using pose-based discriminant embedding,” Signal Process-Image Commun., vol. 27, no.1, pp. 96-111, Jan. 2012.

[15] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J.

Garcia-Rodriguez, “A survey on deep learning techniques for image and video semantic segmentation,” Appl. Soft Comput., vol. 70, pp. 41-65, Sep. 2018.

[16] J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural Netw., vol. 61, pp. 85-117, Jan. 2015.

[17] S. Ji,W. Xu, M. Yang, and K.Yu, “3D Convolutional Neural Networks for Human Action Recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 1, pp. 221-231, Mar. 2012.

[18] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos, ” in Proc. 2014 IEEE Int. Conf. Neural. Inf. Process. Syst., pp. 568-576, Kuching, Malaysia, Nov. 2014.

[19] C. Feichtenhpfer, A. Pinz, and A. Zisserman, “Convolutional twostream network fusion for video action recognition,” in Proc. 2016 Int. Conf. Comput. Vis. Pattern Recognit., pp. 1933-1941, Las Vegas, USA, Jun. 2016.

[20] Harikrishna. B, Ramakrishnam Raju S.V.S, Sanjay Kumar Suman, Bhagyalakshmi L, “Artificial Intelligence Framework for Rice Blast Disease Detection and Classification Using Recurrent Neural Networks”, IET-IEEE 8th International Conference on Computing in Engineering and Technology (ICCET 2023), Patna 14-15, July 2023. doi: https://doi.org/10.1049/icp.2023.1461.

[21] Ramakrishnam Raju S.V.S, Harikrishna. B, Bhagyalakshmi. L, Sanjay Kumar Suman, “Anomaly detection using self-supervised label generator with vertical disparity maps”, IET-IEEE 8th International Conference on Computing in Engineering and Technology (ICCET 2023), Patna 14-15, July 2023. doi: https://doi.org/10.1049/icp.2023.1461.

[22] M.D. Javeeda, Regonda Nagarajub, Raja Chandrasekaran, Govinda Rajulu, Praveen Tumuluru, M. Ramesh, Sanjay Kumar Suman, and Rajeev Shrivastava, “Brain tumor segmentation and classification with hybrid clustering, probabilistic neural networks”, Journal of Intelligent & Fuzzy Systems, IoS Press, vol. 45, issue 4, pp. 6485-6500, 2023. DOI: http://doi.org/10.3233/JIFS-232493.

[23] Himanshu Shekhar, Chandra Bhushan Mahato, Sanjay Kumar Suman, Satyanand Singh, L. Bhagyalakshmi, Mahendra Prasad Sharma, B. Laxmi Kantha, Helan Vidhya T, Siva Kumar Agraharam & A. Rajaram, “Demand Side Control for Energy Saving in Renewable Energy Resources Using Deep Learning Optimization,” Electric Power Components and Systems, Taylor & Francis, vol. 51, issue 19, pp. 2397-2413, 2023. DOI: http://doi. org/10.1080/15325008.2023.2246463.

[24] Sanjay Kumar Suman, Himanshu Shekhar, Chandra Bhushan Mahto, D. Gururaj, L. Bhagyalakshmi, and P. Santosh Kumar Patra, “Sign Language Interpreter”, Springer Nature Singapore, Advances in Cognitive Science and Communications, Cognitive Science and Technology, pp. 1021-1031, 2023, https://doi.org/10. 1007/978-981-19-8086-2_96

[25] Rajesh Tiwari, Rajeev Shrivastava, Santosh Kumar Vishwakarma, Sanjay Kumar Suman, and Sheo Kumar, “InterCloud: Utility-Oriented Federation of Cloud Computing Environments Through Different Application Services”, Springer Nature Singapore, Advances in Cognitive Science and Communications, Cognitive Science and Technology, pp. 83-89, 2023. https://doi.org/10. 1007/978-981-19-8086-2_8

[26] Rajeev Shrivastava, Mallika Jain, Santosh Kumar Vishwakarma, L. Bhagyalakshmi, and Rajesh Tiwari, “Cross-Cultural Translation Studies in the Context of Artificial Intelligence: Challenges and Strategies”, Springer Nature Singapore, Advances in Cognitive Science and Communications, Cognitive Science and Technology, pp. 91-98, 2023. https://doi.org/10.1007/978-981-19-8086-2_9

[27] Abhishek Kumar, Sanjay Kumar Suman, L. Bhagyalakshmi, and Anil Kumar Sahu, “IoT and Cloud Network Based Water Quality Monitoring System Using IFTTT Framework”, Springer Nature, Sustainable Technology and Advanced Computing in Electrical Engineering, Lecture Notes in Electrical Engineering 939, pp. 23-32, 2023. https://doi.org/10.1007/978-981-19-4364-5_3

[28] Sanjay Kumar Suman, Dhananjay Kumar and L. Bhagyalakshmi, “SINR pricing in non- cooperative power control game for wireless ad hoc network”, KSII Transactions on Internet and Information Systems, KSII TIIS, vol. 8, no. 7, pp. 2281-2301, 2014. https://dio. org/10.3837/tiis.2014.07.005

[29] L. Bhagyalakshmi, Sanjay Kumar Suman, Sujeetha Devi, “Joint Routing and Resource Allocation for Cluster Based Isolated Nodes in Cognitive Radio Wireless Sensor Networks”, Wireless Personal Communication, Springer, vol. 114, issue 4, pp. 3477- 3488, Oct. 2020. https://doi.org/10.1007/s11277-020-07543-4

[30] K. Mahalakshmi, K. Kousalya, Himanshu Shekhar,

Aby K. Thomas, L. Bhagyalakshmi, Sanjay Kumar Suman et. al., “Public Auditing Scheme for Integrity Verification in Distributed Cloud Storage System”, Scientific Programming, Hindawi, vol. 2021, Article ID 8533995, Dec. 2021. .https://doi.org/10.1155/2021/8533995

[31] Sanjay Kumar Suman et. al., Detection and prediction of HMS from drinking water by analysing the adsorbents from residuals using deep learning, Hindawi (SAGE Journal) Adsorption Science & Technology, vol. 2022, Article id 3265366, March 2022. https://doi.org/10.1155/2022/3265366

[32] Bhagyalakshmi and K. Murugan, “Avoiding Energy Holes Problem using Load Balancing Approach in Wireless Sensor Network”, KSII Transaction on Internet and Information Systems, vol. 8 , no. 5, pp. 1618- 1637, 2014. https://dio.org/10.3837/tiis.2014.05.007.

[33] Satyanand Singh, Sajai Vir Singh, Dinesh Yadav, Sanjay Kumar Suman, Bhagyalakshmi Lakshminarayanan, Ghanshyam Singh, “Discrete interferences optimum beamformer in correlated signal and interfering noise, International Journal of Electrical and Computer Engineering, vol. 12, no. 2, pp. 1732-1743, April 2022. http://doi.org/10.11591/ijece.v12i2.pp1732-1743.