The Scientific Temper (2023) Vol. 14 (4): 1185-1191
Doi: 10.58414/SCIENTIFICTEMPER.2023.14.4.19

E-ISSN: 2231-6396, ISSN: 0976-8653
https://scientifictemper.com/

RESEARCH ARTICLE

Efficient I3D-VGG19-based architecture for human activity recognition

Rashmika Vaghela[1*], Dileep Labana[1], Kirit Modi[2]

Abstract

Systems for autonomously identifying and analyzing human human activity recognition (HAR) using video data captured from various devices require ongoing updates due to HAR's developing technology and multidisciplinary nature. This research aims to classify different activities performed by humans using various pre-trained models and latest transfer learning methods. Set the hyperparameter values to get accurate classification based on different performance evaluation matrices. In this study, the VGG19 based optimized I3D architecture is proposed. The experimental findings demonstrate that use of an optimized VGG19 based I3D model on the UCF-50 dataset has led to an enhancement in the performance of the human activity recognition system with an accuracy rate of training is 98.24% and testing is 98.36%, surpassing the performance of alternative I3D model using DenseNet121 in direct comparison. This will facilitate the development of applications like smart environments, elderly care and assistive technologies, healthcare and wellness, and other domains.

Keywords: Human activity recognition, DenseNet121, VGG19, Convolutional neural networks.

Introduction

One of the most interesting and useful areas of research in computer vision is automatic human activity identification. In these systems, the appearance and patterns of the motions in the video sequences serve as the basis for human activity labeling; however, most of the existing research, most conventional methodologies, and classic neural networks either ignore or are unable to use temporal information for activity recognition prediction from a video sequence. However, proper and accurate human activity recognition (HAR) comes at a hefty computational cost (Serpush, F. & Rezaei, M., 2021). Convolutional neural networks (CNNs) are increasingly being used as a feature learning technique for HAR. Features may be automatically extracted by CNNs. CNNs, on the other hand, need a training phase, which makes them vulnerable to the cold-start issue (Cruciani *et al*., 2020). The CNN architecture has been producing improved outcomes for image datasets because of its capacity to learn and extract the important information from the images (Savyanavar, A.S., Mhala, N.C., & Sutar, S.H., 2023). Traditional machine learning techniques for recognizing human activities extracted visual information mostly from human observations. It necessitates extensive human experience and expertise. The majority of these algorithms were only effective on the precise dataset used in a particular experiment. The Internet is currently home to many digital video materials (Yu, Z., & Yan, W. Q., 2020, November 1). The utilization of deep learning (DL) in recent studies has demonstrated enhanced accuracy compared to conventional machine learning (ML) algorithms (Math, S., 2022). Artificial neural network-based feature maps are produced using deep learning algorithms. In the areas of robotics, natural language processing, and computer vision, deep neural networks have made outstanding advancements (Chai *et al*., 2021). Human motion is rather complex, and several factors, such a chaotic backdrop, different lighting conditions, unreliable picture capture, a lack of suitable pattern classes, etc., might impact the relevant motion analysis. As a result, deep learning has plenty of potential to expand human action detection. Deep learning is also incredibly important for implementing self-learning and transfer learning (Vaghela, R. K., Patel, J. A., & Modi, K., 2022).

[1]Department of Computer Science & Engineering, Parul University Vadodara, Gujarat, India.

[2]Department of Computer Engineering, Sankalchand Patel University, Visnagar, India.

***Corresponding Author:** Rashmika Vaghela, Department of Computer Science & Engineering, Parul University Vadodara, Gujarat, India, E-Mail: rashmivaghela.rv@gmail.com

How to cite this article: Vaghela, R., Labana, D., Modi, K. (2023). Efficient I3D-VGG19-based architecture for human activity recognition. The Scientific Temper, **14**(4):1185-1191.

Doi: 10.58414/SCIENTIFICTEMPER.2023.14.4.19

Source of support: Nil

Conflict of interest: None.

Received: 01/10/2023 **Accepted:** 02/11/2023 **Published** : 27/12/2023

In this paper, we proposed an optimized I3D-VGG19 pre-trained deep learning model-based (H. Mohamed, E., H. El-Behaidy, W., Khoriba, G., & Li, J., 2020) architecture for HAR system which recognizes human activities from video input files. The significance of the proposed model is given below:

- To improve the performance of human activity classification, it is suggested that the pre-trained 2D ConvNet be inflated by adding an extra dimension for the temporal axis.
- The suggested approach takes use of pretrained VGG19 weights with optimization using the ImageNet dataset.
- It is suggested that use ReLU activation function to enhance the performance of the human activity recognition system.
- The proposed approach successfully classifies human activities from video input with astounding accuracy.

After the introductory section, the subsequent sections of this paper are structured in the following manner: Section II provides an overview of the existing literature and research that is relevant to the topic at hand. Section III of this study pertains to the methodology and proposed system. Section IV of the document presents the details of the results. Following this, section V provides a comprehensive conclusion.

Literature Survey

Human activity recognition may be accomplished using a variety of methods, including computer vision, deep learning, sensor-based, machine learning, and sensor-based methods (Gupta *et al.*, 2022).

Three algorithms are based on CNN, two-stream CNN, CNN+LSTM, and 3D CNN to identify human video actions. Designed and implemented in (Yu, Z., & Yan, W. Q., 2020, November 1). Each algorithm was explicated and analyzed on detail. HMDB-51 dataset was applied to test these algorithms and gain the best results. Experimental results showcase that the three methods have effectively identified human actions given a video. The best algorithm thus was selected.

CNN-GRU Model for the prediction of HAR has been proposed in (Math, S., 2022). They used the UCI HAR dataset and achieved 96.79% accuracy for six normal day-day activities.

A novel structure named hierarchical deep long short term memory (H-LSTM) based on long short-term memory has been developed in (Wang, L., & Liu, R., 2019). Smoothing and denoising the raw sensor data prepared it for selection and feature extraction using the time-frequency-domain approach. Second, the categorization of these activities is done using H-LSTM. Three publicly available UCI datasets are employed to simulate the automated feature vector extraction and categorization of output recognition results. The simulation findings also show that the H-LSTM network outperforms other deep learning techniques. Human activity identification using the H-LSTM network has been shown to be 99.15% accurate.

A novel deep neural network for recognizing human behaviors that blends LSTM and convolutional layers has been proposed (Xia, K., Huang, J., & Wang, H., 2020). The CNN weight settings were largely focused on the fully connected layer. In response to this trait, a GAP layer was used in place of the fully connected layer beneath the convolutional layer, substantially lowering the model parameters while retaining a high recognition rate. A BN layer was also added after the GAP layer to hasten the model's convergence, and a clear result was achieved. In the proposed architecture, the raw data collected by mobile sensors were fed into a two-layer LSTM, followed by convolutional layers, allowing it to learn the temporal dynamics on various time scales in line with the learned LSTM parameters for increased accuracy. The experiment used the three available datasets UC-HAR, WISDM, and OPPORTUNITY to show the suggested model's effectiveness. Because accuracy was not a sufficient and complete measure of performance, the F1 score was used to evaluate the model's performance. The F1 score was finally 95.78, 95.85, and 92.63% on the UCI-HAR, WISDM, and OPPORTUNITY datasets. Additionally, they looked at how different hyper-parameters, such as batch size, optimizer type, and number of filters, affected model performance.

1D-CNN, LSTM, and bi-directional LSTM have been proposed in (Fard Moshiri *et al.*, 2021). The Raspberry Pi 4 was employed. The CSI data was then converted into pictures, and the generated images served as the inputs for a classifier powered by a 2D CNN. In tests, it was shown that the recommended CSI-based HAR outperformed competitors' methods and had an accuracy of nearly 95% for seven daily tasks.

Different machine learning algorithms were implemented in (Gutiérrez-Esparza *et al.*, 2021). According to the results, the random forest obtained the highest performance in identifying the best features. A summary of the literature review is shown in Table 1.

There hasn't been much research on video files for identifying human activities. Most of the work is done on the sensor-based, human posture, and digital picture datasets. If a wearable sensor-based approach is considered to be an effective HAR solution. The low recognition rate may be caused by a number of things, including improper placement or orientation of sensors, managing the massive amounts of data that the devices can generate, managing their temporal dependency, and, second, a lack of understanding of how to relate this data to the defined movements.

The findings of this research show that in these circumstances, utilizing a pose dataset technique yields better results; nevertheless, there is no online posture dataset that is adequate for training or classification tasks.

Table 1: Summary of related research work

Researcher	*Algorithm*	*Dataset*	*Accuracy (%)*
Z., & Yan, W. Q.	Two-stream CNN, CNN+LSTM, and 3D CNN	HMDB-51	Mentioned that they got best result
Math, S.	CNN+GRU	UCI-HAR	96.79
Wang, L., & Liu, R.	deep LSTM	UCI	99.15
Xia, K., Huang, J., & Wang, H.	CNN+LSTM	UC-HAR,	95.78
		WISDM,	95.85
		OPPORTUNITY	92.63,
Fard Moshiri *et al.*	1D-CNN, LSTM, and bi-directional LSTM	Seven daily activity image dataset	95

A single photograph cannot adequately capture all of the activities that people participate in, despite the fact that we can classify human activity using an image collection. The greatest choice for human activity detection will be a video processing-based approach since it will provide better results with more powerful computers. For precise activity detection, pre-activity and post-activity data are also crucial (temporal).

Methods

This section describes the proposed I3D system using optimized VGG19 model. Here we used DenseNet121 based I3D system, to compare the results with the proposed optimized VGG19 based I3D model. Working of HAR system has been shown in Figure 1.

UCF-50 Dataset

The UCF-50 dataset was used in the research (Reddy, K. K., & Shah, M., 2012). The dataset consists of video files. The dataset contains a comprehensive collection of video files related to activities performed by human, encompassing 50 distinct classes. The collection has 25 sets of films, one for each class type.

Data Pre-Processing

Following the acquisition of the dataset, we pre-processed videos of the dataset extract frames, resized and normalized frames from videos. Generate frame and label list for later analysis and processing. This study considers 64 x 64 pixels values for images. Normalization reduces computational complexity during the training the model. However, using Eq. (1) images were normalized.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

In order to mitigate the issue of overfitting, a portion of the training data, specifically 25%, is allocated for the purpose of validations.

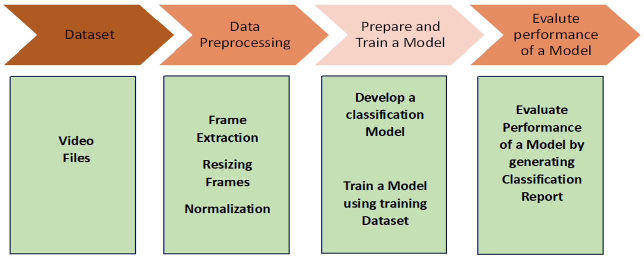

Dataset
Data Preprocessing
Prepare and Train a Model
Evalute performance of a Model
Video Files
Frame Extraction
Resizing Frames
Normalization
Develop a classification Model
Train a Model using training Dataset
Evaluate Performance of a Model by generating Classification Report

Figure 1: working architecture of human activity recognition system

Feature Extraction and Classification

For feature extraction and classification from the frames, here we used I3D architecture. We used pre-trained CNN model VGG19 and DenseNet121.

I3D, short for Inflated 3D ConvNet, is a popular architecture used for video action recognition and understanding. I3D is known for its effectiveness in capturing spatiotemporal information from video data makes it a powerful tool for recognizing video actions and activities (Gowada, R., Pawar, D., & Barman, B., 2023).

Details about I3D Model

Spatiotemporal convolution

i3D is an extension of the inception architecture, which is typically used for image classification. The innovation in I3D includes 3D convolutions, which can capture both spatial and temporal features in video data. These 3D convolutions are used to process video frames over time, effectively modeling the motion and appearance of objects in the video.

Inflated convolution

The term “inflated” in I3D refers to the process of converting 2D convolutions into 3D convolutions. This is done by initializing the 3D convolution kernels with 2D kernels and leaving the third dimension (time) mostly untouched. This allows the model to take advantage of pre-trained 2D CNNs for image recognition, such as inception or ResNet, and adapt them for video analysis.

DenseNet121

A deep neural network design called densely connected convolutional networks (DenseNet) incorporates the idea of dense connections between layers. The difficulties of gradient vanishing and feature reuse can arise in conventional CNNs are addressed by DenseNet designs. Variation of the DenseNet architecture is represented by DenseNet121 as shown in Figure 2.

VGG19

For picture classification and object identification applications, the deep convolutional neural network architecture VGG19 was developed as shown in Figure 3. It is a member of the visual geometry group (VGG) series of architectural designs created by Oxford University academics. Convolutional and completely linked layers are among the 19 layers that make up the VGG19 design, which is an expansion of the original VGG16 structure.

VGG19 layer architecture and its components

- *Input layer*

VGG19 takes an input image of size (224, 224, 3), where 224x224 is the spatial resolution of the image, and 3 represents the number of color channels (RGB).

- *Convolutional layers*

VGG19 consists of a series of convolutional layers, each followed by a rectified linear unit (ReLU) activation function and a max-pooling layer. The convolutional layers use small receptive fields (3x3) and a stride of 1, which helps capture fine details in the images.

- *Max-Pooling layers*

After each group of convolutional layers, max-pooling layers are applied to down sample the feature maps and reduce the spatial dimensions. The max-pooling layers use a 2x2 window with a stride of 2, halving the dimensions of the feature maps.

- *Fully connected layers*

VGG19 includes a series of fully connected layers for classification after the convolutional and pooling layers. ReLU activation functions follow the fully connected layers.

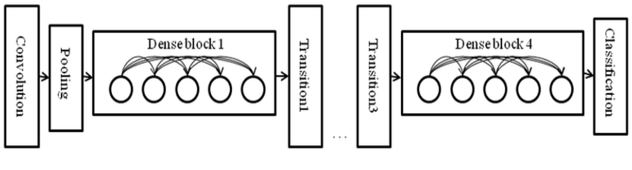


Figure 2: Layer architecture of DenseNet121

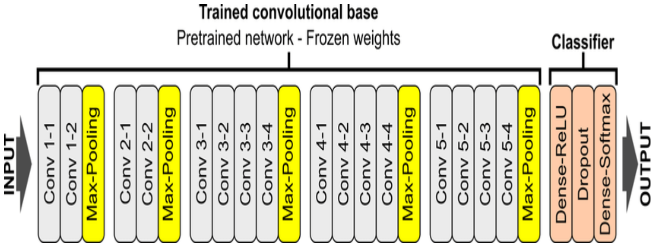


Figure 3: Layer architecture of VGG-19

Table 2: Layers of I3D-Densenet121 model

Layer (type)	*Output shape*	*Param #*
input_2 (InputLayer)	[(None, 30, 64, 64, 3)]	0
time_distributed (TimeDistributed)	(None, 30, 2, 2, 1024)	7037504
time_distributed_1 (TimeDistributed)	(None, 30, 1024)	0
dropout (Dropout)	(None, 30, 1024)	0
time_distributed_2 (TimeDistributed)	(None, 30, 512)	524800
dropout_1 (Dropout)	(None, 30, 512)	0
time_distributed_3 (TimeDistributed)	(None, 30, 5)	2565

Total params: 7,564,869
Trainable params: 527,365
Non-trainable params: 7,037,504

Table 3: Layers of I3D- VGG-19 model

Layer (type)	*Output shape*	*Param #*
input_2 (InputLayer)	[(None, 30, 64, 64, 3)]	0
time_distributed (TimeDistributed)	(None, 30, 2, 2, 512)	20024384
time_distributed_1 (TimeDistributed)	(None, 30, 512)	0
dropout (Dropout)	(None, 30, 512)	0
time_distributed_2 (TimeDistributed)	(None, 30, 512)	262656
dropout_1 (Dropout)	(None, 30, 512)	0
time_distributed_3 (TimeDistributed)	(None, 30, 5)	2565

Total params: 20,289,605
Trainable params: 265,221
Non-trainable params: 20,024,384

- *Flatten layer*

Before the fully connected layers, a flatten layer is applied to convert the 3D feature maps into a 1D vector, which serves as the input to the fully connected layers.

- *Output layer*

The final fully connected layer has as many neurons as the number of classes in the classification task.

A softmax activation function is applied to the output layer to convert the network's raw scores into class probabilities.

The layer structure of I3D model with pre-trained DenseNet121 is shown in Table 2.

Layer structure of I3D model with pre-trained optimized VGG-19 is as shown in Table 3.

For model training, we utilize the following parameters:
Epoch: 50
Learning rate:0.0001
Optimizer: Nadam
Batch size: 20
We are using early stopping callback

To enhance our model performance, we use rectified linear unit (ReLU) activation function (Agarap, Abien Fred., 2018). Non-linearity is introduced through the ReLU activation function. The rectifier or ReLU activation function is an activation function defined in Eq. (2).

$$f(x) = x^{+} = \max(0, x) = \frac{x + |x|}{2} = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \qquad (2)$$

Where x is the input to a neuron.

Performance assessment

To check the effectiveness of the classification system a confusion matrix that includes important metrics like true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) are used. These characteristics may be used to calculate validity metrics like precision, F1-score, accuracy, and recall. Formulas for the validity metrices are mentioned in Eq. (3), (4), (5) and (6).

Accuracy = (TP+TN)/ (TP+ FP + TN + FN) (3)

Precision = TP / (TP + FP) (4)

Recall = TP / (TF + FN) (5)

$$\text{F1 Score} = 2^{*} \frac{\text{(Precision x Recall)}}{\text{(Precision + Recall)}} \qquad (6)$$

Results

We use Kaggle.com, a website made specifically for deep learning applications, to implement the experiment for this study. The current study uses a GPU T4 X 2 that can

Table 4: Classification report of I3D model with VGG19

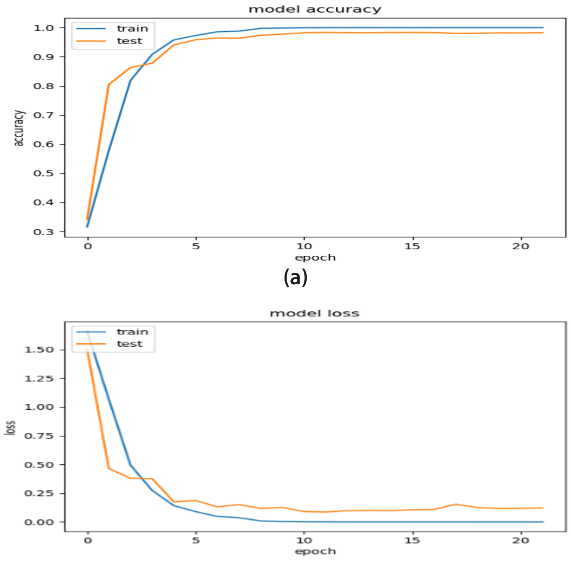
Activity class	*Precision*	*Recall*	*F1-score*
Horse race	1.00	0.99	0.99
Kayaking	0.98	0.99	0.99
Swing	0.99	1.00	0.99
TaiChi	0.96	1.00	0.98
Walking with dog	0.97	0.93	0.95

Table 5: Classification report of I3D model with Densenet121

Activity class	*Precision*	*Recall*	*F1-score*
Horse race	0.98	0.99	0.99
Kayaking	0.95	1.00	0.97
Swing	0.99	0.98	0.97
Taichi	0.96	0.97	0.98
Walking with dog	0.97	0.90	0.93

Table 6: Comparison of both I3D model results

Method	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1-score (%)*
I3D model with VGG19	98.36	98.00	98.00	98.00
I3D model with DenseNet121	96.31	96.00	96.00	96.00



(a)

(b)

Figure 4: (a)Accuracy and (b) Loss graph of I3D model with VGG19

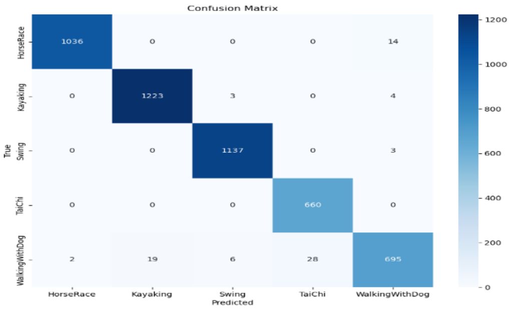
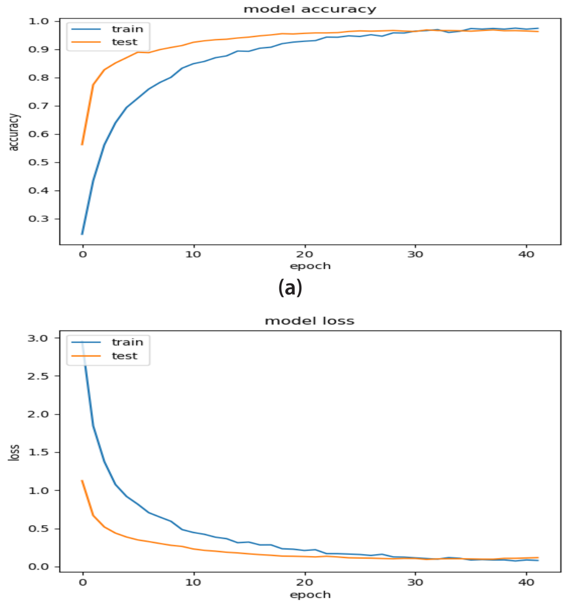


Figure 5: Confusion matrix for optimized I3D-VGG19 Model



(a)

(b)

Figure 6: (a)Accuracy and (b) Loss graph of I3D model with DenseNet121

be accessed *via* a gateway to enhance and speed up the computing operations. This experiment was carried out using the Keras and Tensorflow libraries.

Performance metrics included for the study include precision, recall, and F1 score. The higher recall score indicates that the model can recognize more true positives reliably. A higher f1-score shows increased model performance.

The training accuracy was determined to be 98.24%. The testing accuracy achieved a value of 98.36% when evaluated on the testing dataset with I3D model with pre-trained VGG19. Accuracy- Loss graphs for I3D model with VGG19 are shown in Figure 4. A classification report and confusion matrix are generated for performance evaluation, as shown in Table 4 and Figure 5.

The training accuracy was determined to be 96.31%. The testing accuracy achieved a value of 96.31% when evaluated on the testing dataset with I3D model with pre-trained DenseNet-121. Accuracy- Loss graphs for I3D model with DenseNet-121 are shown in Figure 6. For performance evaluation classification report and confusion matrix are generated, as shown in Table 5 and Figure 7.

Table 6 presents a comparative analysis, showcasing the results of both models on the UCF-50 dataset for human activity classification.

Results Comparison

In Figure 8, an examination that compares the outcomes of our model under discussion with earlier studies on the identification of human activities is offered below.

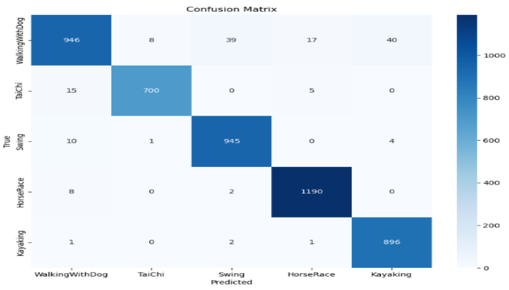


Figure 7: Confusion matrix for optimized I3D-DenseNet121 model

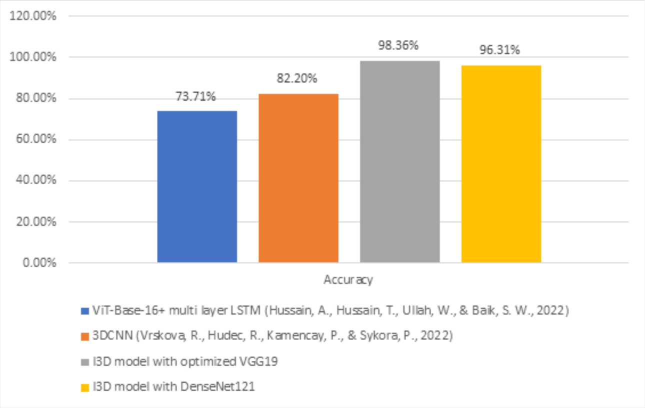


Figure 8: Comparison graph of proposed results with past research work

Discussion

We compared our findings with earlier work in order to effectively and accurately tackle the challenge of human activity detection from video input dataset.

The spatial characteristics were extracted using a pretrained ViT - Base-16 model at predetermined time stamps by Hussain, A., Hussain, T., Ullah, W., and Baik, S. W. (2022). The multilayered LSTM network is then given this spatial information to learn temporal relationships. They carried out in-depth tests on the widely used HAR dataset, UCF-50, and got an accuracy of 73.71%.

The 3D convolutional neural network (3DCNN) was proposed by Vrskova, Hudec, Kamencay, and Sykora (2022) for the identification of human activity in video data. They conducted an experiment using the UCF-50 standard HAR dataset and obtained an accuracy of 82.2%.

We evaluated our I3D-VGG19 optimized model on the same dataset as earlier studies and found that it performed with an accuracy of 98.24% during model training and 98.36% during model testing.

Conclusion

In order to recognize human activity, this study offers two I3D models with pre-trained DenseNet-121 and optimized VGG19 on a UCF-50 dataset. Classification accuracy increased significantly with the introduction of improved VGG19, rising to 98.36%. I3D model with VGG19 performs better than I3D model with DenseNet121.

The findings indicate that the recommended model is the most effective model for detecting human activity based on video input. Its excellent accuracy and low loss point to its resilience in identifying and categorising the wide range of human activities included in the dataset. The proposed paradigm works directly with video files. Processing power, deployment constraints, and the desired balance of accuracy and efficiency can all have an impact on whatever model is used.

Acknowledgement

The authors are highly thankful to Parul University, Vadodara, Gujarat for providing the necessary facilities for this research work.

References

Agarap, A. F. (2018). Deep Learning using Rectified Linear Units (ReLU). http://arxiv.org/abs/1803.08375

Chai, J., Zeng, H., Li, A., & Ngai, E. W. T. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. Machine Learning with Applications, 6, 100134. https://doi.org/10.1016/j.mlwa.2021.100134

Cruciani, F., Vafeiadis, A., Nugent, C., Cleland, I., McCullagh, P., Votis, K., Giakoumis, D., Tzovaras, D., Chen, L., & Hamzaoui, R. (2020). Feature learning for Human Activity Recognition using Convolutional Neural Networks: A case study for Inertial Measurement Unit and audio data. CCF Transactions on Pervasive Computing and Interaction, 2(1), 18–32. https://doi.org/10.1007/s42486-020-00026-2

Fard Moshiri, P., Shahbazian, R., Nabati, M., & Ghorashi, S. A. (2021). A CSI-Based Human Activity Recognition Using Deep Learning. Sensors (Basel, Switzerland), 21(21). https://doi.org/10.3390/s21217225

Gowada, R., Pawar, D., & Barman, B. (2023). Unethical human action recognition using deep learning-based hybrid model for video forensics. Multimedia Tools and Applications, 82(19), 28713–28738. https://doi.org/10.1007/s11042-023-14508-9

Gupta, N., Gupta, S. K., Pathak, R. K., Jain, V., Rashidi, P., & Suri, J. S. (2022). Human activity recognition in artificial intelligence framework: a narrative review. Artificial Intelligence Review, 55(6), 4755–4808. https://doi.org/10.1007/s10462-021-10116-x

Gutiérrez-Esparza, G. O., Ramírez-Delreal, T. A., Martínez-García, M., Infante Vázquez, O., Vallejo, M., & Hernández-Torruco, J. (2021). Machine and deep learning applied to predict metabolic syndrome without a blood screening. Applied Sciences (Switzerland), 11(10). https://doi.org/10.3390/app11104334

Hussain, A., Hussain, T., Ullah, W., & Baik, S. W. (2022). Vision Transformer and Deep Sequence Learning for Human Activity Recognition in Surveillance Videos. Computational Intelligence and Neuroscience, 2022. https://doi.org/10.1155/2022/3454167

Math, S. (2022). Hybrid deep learning framework for human activity recognition. International Journal of Nonlinear Analysis and Applications

Mohamed, E. H., El-Behaidy, W. H., Khoriba, G., & Li, J. (2020). Improved white blood cells classification based on pre-trained deep learning models. Journal of Communications Software and Systems, 16(1), 37–45. https://doi.org/10.24138/jcomss.v16i1.818

Reddy, K. K., & Shah, M. (2013). Recognizing 50 human action categories of web videos. Machine Vision and Applications, 24(5), 971–981. https://doi.org/10.1007/s00138-012-0450-4

Sadanand Savyanavar, A., Mhala, N., & Sutar, S. H. (2023.). Star-galaxy classification using machine learning algorithms and deep learning. In International Journal on Information Technologies & Security, № (Vol. 2).

Serpush, F., & Rezaei, M. (2021). Complex Human Action Recognition Using a Hierarchical Feature Reduction and Deep Learning-Based Method. SN Computer Science, 2(2). https://doi.org/10.1007/s42979-021-00484-0

Vaghela, R. K., Patel, J. A., & Modi, K. (2022). Human Activity Recognition Using Feature Fusion. SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology, 14(2), 288–293. https://doi.org/10.18090/samriddhi.v14spli02.25

Vrskova, R., Hudec, R., Kamencay, P., & Sykora, P. (2022). Human Activity Classification Using the 3DCNN Architecture. Applied Sciences (Switzerland), 12(2). https://doi.org/10.3390/app12020931

Wang, L. K., & Liu, R. Y. (2020). Human Activity Recognition Based on Wearable Sensor Using Hierarchical Deep LSTM Networks. Circuits, Systems, and Signal Processing, 39(2), 837–856. https://doi.org/10.1007/s00034-019-01116-y

Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN Architecture for Human Activity Recognition. IEEE Access, 8, 56855–56866. https://doi.org/10.1109/ACCESS.2020.2982225

Yu, Z., & Yan, W. Q. (2020). Human Action Recognition Using Deep Learning Methods. International Conference Image and Vision Computing New Zealand, 2020-November. https://doi.org/10.1109/IVCNZ51579.2020.9290594