HUMAN ACTIVITY RECOGNITION USING K-NEAREST NEIGHBOUR MACHINE LEARNING ALGORITHM

[1] **Doppala sravani,** Pg Scholar, Department of E.C.E, VIKAS GROUP OF INSTITUTIONS, Email id: sravaniyashoda817@gmail.com.
[2] **Yarram Uma Maheswari,** Assistant professor, Department of E.C.E, VIKAS GROUP OF INSTITUTIONS, Email id: umaecestaff@gmail.com.
[3] **V L N Phani Ponnapalli,** Associate Professor, Department of E.C.E, VIKAS COLLEGE OF ENGINEERING AND TECHNOLOGY, Email id: pvlnphani0454@gmail.com.

Abstract
Human Activity recognition is now an important concept in many applications. Machine learning (ML) algorithms can help precisely recognize the human activities, provided that well-designed and trained ML algorithms for high performance recognition are developed. This paper presents a k-nearest neighbor (KNN) algorithm for classification of human activities, namely jogging, walking, boxing, hand waving, cycling and surfing. This algorithm is trained and the algorithm's parameters are precisely tuned of for high accuracy achievement. Experimentally, a normalized confusion matrix, a classification report of human activities, receiver operating characteristic (ROC) curves, and precision-recall curves are used to analyze the performance of the KNN algorithm. The results show that the KNN algorithm provides a high performance in the classification of human activities. The weighted average precision and the area under the micro-average precision-recall curve for the KNN more respectively.

Keywords: Machine learning, KNN, Human activity recognition.

1 Introduction

Due to development on laptop vision, computer systems enhance at the decision of a few very tough problems (which includes information an photograph). Models are made wherein the version can are expecting what the photograph is or can locate whether or not or now no longer a selected item is gift withinside the photograph if an photograph is given to the version. These fashions are called neural networks (or synthetic neural networks) stimulated through a human mind shape and function. Deep getting to know, a subfield of system getting to know is the take a look at of those neural networks, which over the years have delivered numerous versions of those networks for diverse problems. For Video Recognition, this method makes use of deep getting to know withinside the context of some of categorized images, a version is constructed in order that it is able to generate a prediction label for a brand new video. Steps had been taken for execution are downloading, extracting and pre-processing a video dataset then dividing the dataset into schooling and checking out information then advent of a neural community and educate it at the schooling information sooner or later checking out the version at the take a look at information [1]. The fourth business revolution (Industry 4.0) is characterised with the aid of using the net of things (IoT), cyber-bodily systems, huge statistics, and synthetic intelligence (AI), which had a essential effect in clever production processes [1]. Moreover, human beings in a clever manufacturing facility being are generally chargeable for the producing processes. So, must be classify human sports to understand the performance of the human beings at some stage in the producing. In different words, with the latest virtual transformation related to the exponential increase of data and communication-pushed technology it's far important to seamlessly figuring out human sports in clever factories [2, 3]. One of the applied gadgets for size and detection of human sports are accelerometers, which can be utilized in hospitals to discover sufferers status [4]. However, to a point commercially to be had accelerometers have constrained accuracy, that's remember an problem to be addressed [5, 6]. More correct popularity of human sports is solved thru a imaginative and prescient technology, together with virtual cameras [7–9], however this approach can not continually attain a first-rate overall performance of accuracy [10, 11]. Machine studying algorithms, together with K-nearest neighbors (KNNS) contributes to a promising strategy to clear up the low accuracy problem. KNN algorithms have already proven excessive feasibility in fields together with facial popularity, textual content mining and doubtlessly relevant for figuring out human sports as well. In the literature, some of gadget studying algorithms were supplied for type of human sports. Chen and Shen [12] brought a random forest (RF) set of rules to categorise human sports, this set of rules used an accelerometer with a gyroscope to seize the statistics, this set of rules become implemented on a dataset together with 27,681 samples and reached a complete accuracy of 83.59%. In [13], Fen et al. proposed a J48 set of rules with a cellphone with integrated accelerometers for day by day lifestyles sports. It become implemented on a dataset carries 8,097 statistics samples and had an accuracy of 89.6%. Ling and Wang [14] supplied a layout of a selection tree (DT) set of rules for human pastime identification, this layout become used with best 4 sports of sitting, walking, standing, and walking and it executed an 73.72% trying out accuracy. Girja et al. [15] brought a naive bayes (NB) set of rules to understand human pastime. This set of rules become additionally implemented to the UCI-HAR dataset [16] and reached an normal accuracy of 79%. In [17], Bao and Intille proposed an set of rules to understand human sports. It used a DT set of rules with five accelerometers. It divided five sports reached an accuracy of 84.0%. Looking the reviewed literature, you can argue that despite the fact that KNN set of rules has proven excessive capability to cope with the accuracy problem in detecting human sports, but they reached overall performance has now no longer been but as much as the bar. In this context, this paper offers an stronger KNN set of rules to categorise human sports, specifically Laying, Downstairs, Sitting, Upstairs, Standing, and Walking, with an goal of growing the type accuracy of the set of rules thru a quality tuning for the parameters of set of rules. This set of rules is evaluated the use of metrics of the precision, F1-score, place below the prec.ision-don't forget curve. The set of rules is implemented to the HAR dataset.

2 Literature Survey

In the subsequent session, we in short assessment exclusive motion popularity strategies. Over the previous few decades, RGB photos has been drastically utilized in human motion popularity strategies. These strategies relies upon on numerous capabilities extracted from the strategies including histogram of orientated gradients (HOG), scale-invariant function transform (SIFT), neighborhood binary styles (LBP), optical float, histogram of float (HOF), regionally binary float styles (LBFP) capabilities as enter to the classifiers including neural community, help vector system (SVM), hidden Markov version (HMM) and different deep mastering strategies [9-11] for human motion popularity. There are nevertheless unsolved troubles in RGB primarily based totally human motion popularity. In specific intensity and skeleton fashions had been proven more popularity charges in comparison to RGB primarily based totally fashions. Recently, with the arrival of low-price sensors including Microsoft Kinect sensors, studies on human motion popularity has shifted to collection of intensity maps and skeleton joints. The benefits of RGB-D sensors over 2D shadeation photo cameras are their invariance to shadeation and lighting fixtures environments.

Using each intensity maps and skeleton joints, D. Thombre, J. Nirmal, and D. Lekha, [1] "Human detection and monitoring the usage of photo segmentation and Kalman filter," Recent advances in photo processing and system mastering strategies have substantially more advantageous the capacity of item category from photos and films in exclusive packages. Classification of human sports is one of the rising studies regions withinside the discipline of laptop imaginative and prescient. It may be utilized in numerous packages which includes clinical informatics, surveillance, human laptop interplay, and mission tracking. In the clinical and healthcare discipline, the category of patients' sports is essential for supplying the specified facts to medical doctors and physicians for remedy reactions and diagnosis. Nowadays, a few studies procedures to understand human interest from films and photos had been proposed the usage of system mastering (ML) and smooth computational algorithms. However, superior laptop imaginative and prescient strategies are nevertheless taken into consideration promising improvement instructions for growing human interest category technique from a chain of video frames. This paper proposes an powerful automatic technique the usage of function fusion and ML strategies. It includes 5 steps, which might be the pre-processing, function extraction, function choice, function fusion, and category steps. Two to be had public benchmark datasets are applied to teach, validate, and check ML classifiers of the advanced technique. The experimental effects of this studies paintings display that the accuracies accomplished are 99.5% and 99.9% on the primary and 2d datasets, respectively. Compared with many present associated procedures, the proposed technique attained excessive overall performance effects in phrases of sensitivity, accuracy, precision, and specificity assessment metric.

M. Sharif, M. A. Khan, T. Akram, M. Y. Javed, T. Saba, and A. Rehman, [2] Human interest tracking withinside the video sequences is an fascinating laptop imaginative and prescient area which includes gigantic packages, e.g., surveillance systems, human-laptop interplay, and site visitors manage systems. In this studies, our number one awareness is in presenting a hybrid method for green category of human sports from a given video collection. The proposed technique integrates 4 main steps: (a) section the shifting gadgets with the aid of using fusing novel uniform segmentation and expectation maximization, (b) extract a brand new set of fused capabilities the usage of neighborhood binary styles with histogram orientated gradient and Harlick capabilities, (c) function choice with the aid of using novel Euclidean distance and joint entropy-PCA-primarily based totally technique, and (d) function category the usage of multi-magnificence help vector system. The 3 benchmark datasets (MIT, CAVIAR, and BMW-10) are used for education the classifier for human category; and for testing, we applied multi-digital digicam pedestrian films together with MSR Action dataset, INRIA, and CASIA dataset. Additionally, the effects also are tested the usage of dataset recorded with the aid of using our studies group. For motion popularity, 4 publicly to be had datasets are decided on including Weizmann, KTH, UIUC, and Muhavi to attain popularity charges of 95.80, 99.30, 99, and 99.40%, respectively, which verify the authenticity of our proposed paintings. Promising effects are accomplished in phrases of more precision in comparison to present strategies.

J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu,[3] This paper offers an automated site visitors surveillance device to estimate essential site visitors parameters from video sequences the usage of handiest one digital digicam. Different from conventional strategies that could classify motors to handiest motors and noncars, the proposed technique has an amazing capacity to categorize motors into extra particular training with the aid of using introducing a brand new "linearity" function in automobile illustration. In addition, the proposed device can nicely address the trouble of automobile occlusions resulting from shadows, which frequently result in the failure of in addition automobile counting and category. This trouble is solved with the aid of using a unique line-primarily based totally shadow set of rules that makes use of a fixed of traces to take away all undesirable shadows. The used traces are devised from the facts of lane-dividing traces. Therefore, an automated scheme to come across lane-dividing traces is likewise proposed. The observed lane-dividing traces also can offer essential facts for function normalization, that could make the automobile length extra invariant, and hence plenty beautify the accuracy of automobile category. Once all capabilities are extracted, an most efficient classifier is then designed to robustly categorize motors into exclusive training. When spotting a automobile, the designed classifier can gather exclusive evidences from its trajectories and the database to make an most efficient selection for automobile category. Since extra evidences are used, extra robustness of category may be accomplished. Experimental effects display that the proposed technique is extra robust, accurate, and effective than different conventional strategies, which make use of handiest the automobile length and a unmarried body for automobile category.

3.Existing System

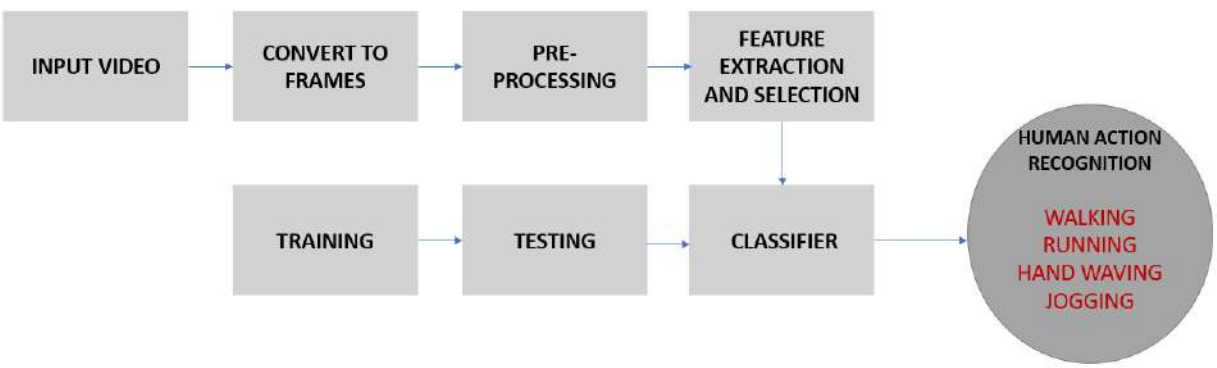

Fig 1 Existing flow chart Human action Recognition

3.1 Dataset pre-processing

The Visual Dataset involves six human activities in four situations-outdoor* s1*, outdoors with [s2*] size, outdoor with specific* s3* clothes and indoor* s4*- (boxing, palpability, hand waving, jogging, jumping, walking), conducted many times in 25 different subjects. Regardless of these circumstances, the layout will be developed. The videos were captured at a frame rate of 25 fps and the resolution of 160x120 pixels was sampled in each frame. The collection comprises 599 images–100 videos in each of the six groups (except Handclapping, which features 99 videos). There are 6 categories- box, handpick, hand-shake, jog and walking [2]. There are a total of 6. When loading the details, we have converted these labels to integer as mentioned below.

ISSN: 0972-2750
Vol-12 Issue-12 No.02: 2022

Action	Coding
Boxing	0
Handclapping	1
Hand waving	2
Jogging	3
Running	4
Walking	5

Table 1. Action Labels Coding

The video's spatial dimension (width x height) can be found to be 160 x 120 pixels. Also, on loading a single video into a NumPy array in python, the shape of the array obtained was (1, 515, 120, 160, 3), this indicates that the video has 515 frames with the spatial dimension of the video is 160 x 120 (width x height) pixels and each frame has 3 channels Red(R), Green(G) and Blue(B).

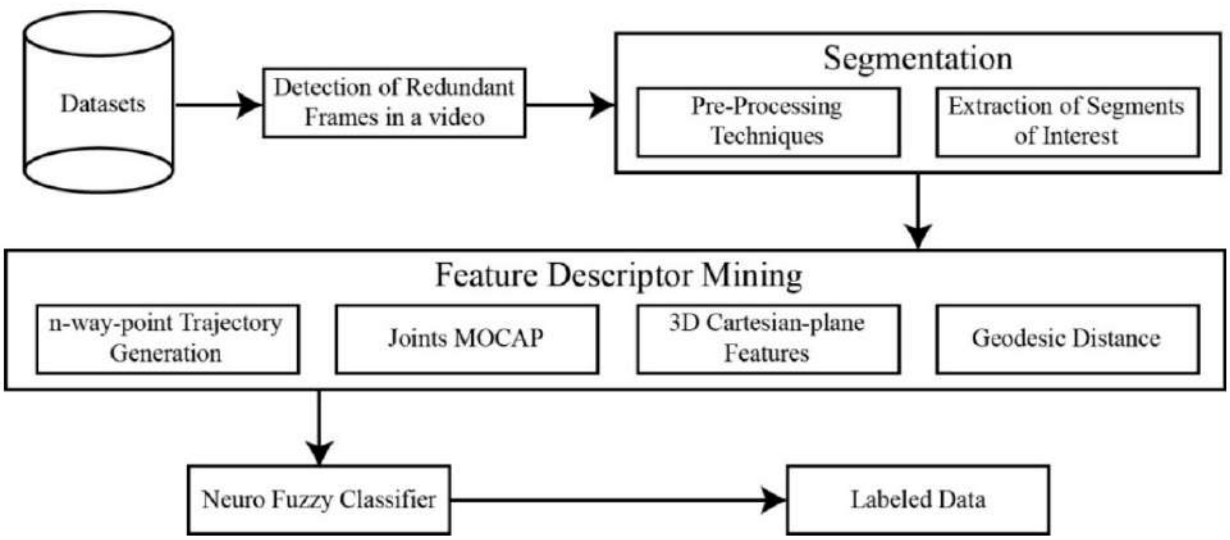

Fig 2 Existing architecture of human action recognition

The video statistics set turned into loaded and the important preprocessing measures had been one of the maximum essential components of the project. Therefore, we created a class (Videos) which turned into renamed (study videos)) (to play and method images. It turned into very hard to create this due to the fact we labored on generalizing this option for any kind of video (now no longer precise to that project). We used NumPy (wherever) to shop and method the videos (with a extra capability tons quicker than the integrated python lists). Figure three exhibits that some of videos of events (no person doing anything), inclusive of the complete guy, may be covered withinside the film. In addition, the bulk of frames could be redundant if the person moved very slowly. The version could face this kind of trouble as a first-rate challenge [3]. The answer over such trouble is received with the aid of using making use of Image AI set of rules which results in kind the frames having human frame found in decided on frame.

Class Label	0	1	2	3	4	5
Boxing	1	0	0	0	0	0
Handclapping	0	1	0	0	0	0
Handwaving	0	0	1	0	0	0
Jogging	0	0	0	1	0	0
Running	0	0	0	0	1	0
Walking	0	0	0	0	0	1

4 Proposed System

In this article, an efficient model for HAR in uncontrolled environment is proposed. The proposed model consists of steps like removing redundant frames from videos, extracting Segments of Interest (SoIs)

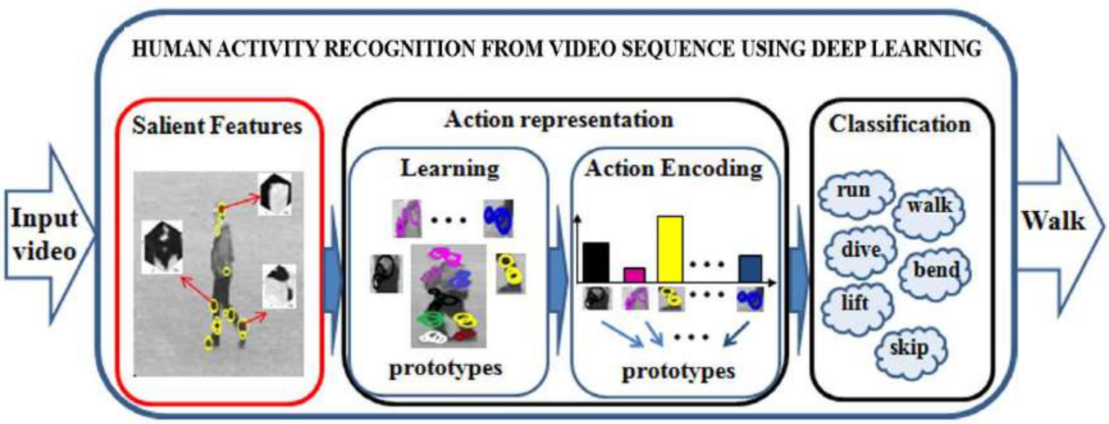

Fig 3 Proposed architecture of Human action recognition

Fig. 3 shows the structure of the proposed KNN algorithm. It is implemented using Scikit-learn framework. It consists of five steps: the first step is a pre-processing using filters to remove the noise from the utilized dataset, the second step is a feature extraction of the dataset. So, 561 features are extracted with fixed windows of 2.56 s “128 readings per window”. The dataset features are ax, ay, and az, which represent the accelerations in the x-axis, y-axis, z-axis, respectively. Also, another extracted features of the dataset are lx, ly, and lz, which represent the gyroscope angles. The third step is the selection of a specific feature, the fourth step is the training on the dataset, and the fifth step is the testing/classification for the activities. Fig. 6.3 illustrates a mechanism of the KNN algorithm. For instance, it assumes two categories or classes “A and B” and a new data point x1 which colored blue. So, the KNN algorithm can detect the data point in which category will be lied. This algorithm uses neighbors’ number of 20 and it is based on the Euclidean distance (dE), which is computed between a two categories “points”, where dE is represented in Eq. (1).

In this algorithm, a neighborhood function is used to classify the activities. The best selection of k depends on the dataset. So, the largest k reduces the noise applied to a classification. Furthermore, the KNN is a supervised machine learning algorithm used to obtain the nearest data point from the same class and nearest data from different classes. The dataset used for the classification is the UCI-HAR dataset available in [16]. It consists of 748,406 samples of different activities Laying, Downstairs, Sitting, Upstairs, Standing, and Walking. The sample percentage for each activity are 18.3%, 14.4%, 16.9%, 15.6%, 18.5%, and 16.3%, respectively. The dataset are gathered from 30 individuals using a mobile phone (Samsung Galaxy S II), which includes inertial sensors positioned in a waist. The dataset readings are extracted with a 50 Hz sampling rate. The dataset has 6 attributes with information correlated to the activities of the humans' activities: time and x-, y-, z-accelerations. The dataset is divided into an 70% training set and a test set of 30%. The test set is utilized to evaluate the proposed algorithm.

Fig. 4 The proposed KNN algorithm structure.

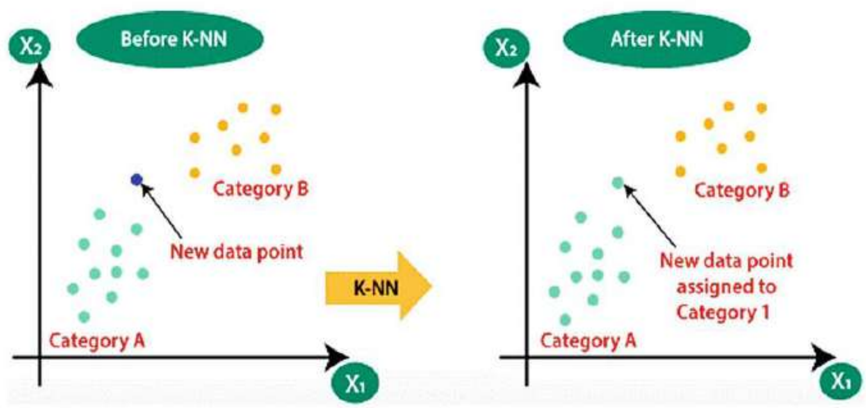

Fig. 5 The mechanism of the KNN algorithm.

$$d_E(x, y) = \sqrt{\sum_{i=1}^{n} (x_1 - y_1)^2}$$

- Reading in the video frame-by-frame.
- The videos were captured at a frame rate of 25fps.
- Each frame needs to have the same spatial dimensions (height and width). Hence each frame in a video will have to be resized to the required size.
- In order to simplify the computations, the frames are converted to grayscale.
- Normalization - The pixel values ranges from 0 to 255. These values would have to be normalized in order to help our model converge faster and get a better performance.

Detecting Redundant Frames

Nowadays, digital cameras collect video data at 60 frames per second (fps) for 4K videos or 30 fps for high-resolution videos to ensure comprehensive smoothness. Surveillance cameras uninterruptedly record videos every time; thus, hundreds of thousands or even millions of video frames are stored every hour. Much computational time and power are required to process these frames efficiently. From these videos' behavior events, it is observed that there exist many static, useless, and redundant frames, which slow down the processing speed of any efficient algorithm. It is necessary to detect and remove redundant frames in long videos to improve algorithms' speed and efficiency. In this article, a method of detecting redundant frames based on Spatiotemporal Interest Points (SIPs) is proposed. Initially, the algorithm calculates the SIPs against each video frame. These interest-points are then combined with temporal and local constraints to detect the static frames. Properties of SIPs manifest that if the quantity and location of interestpoints have not changed in a few of the video frames, it is considered that the content of the subject video has not changed during those frames. This property can be employed by keeping a single frame instead of redundant frames. The proposed algorithm detects interest points at temporal and spatial scales by implementing.

Result

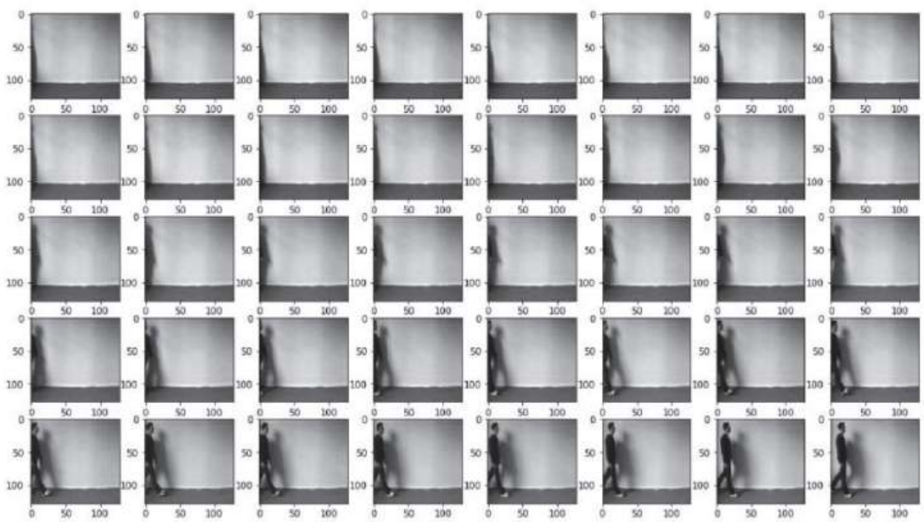

Fig.6 Consecutive frames of a sample video of 'Walking' of KTH dataset

Here in fig 7.1 we had given an video as input in which we had taken 50 frames of the input video and the analysis of the input parameters is done using KNN.

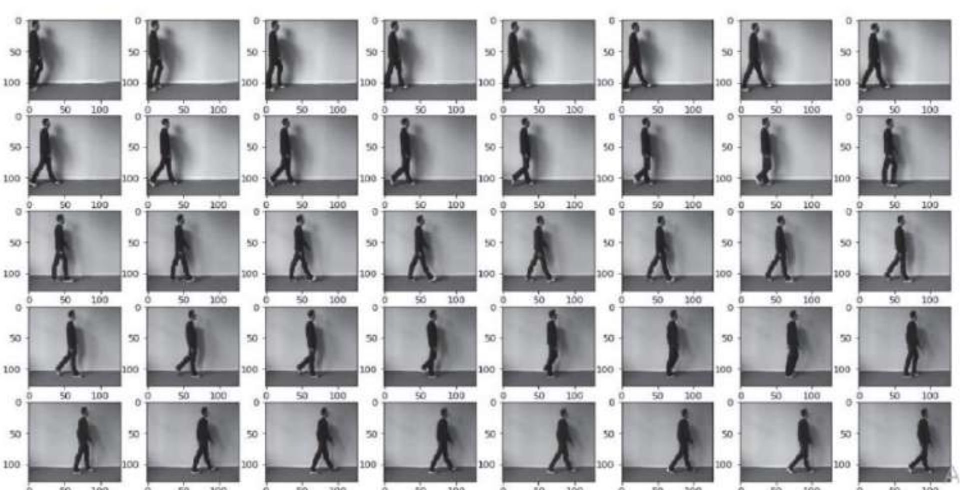

Fig. 7 The frames of a sample video of 'Walking' of KTH dataset with Image AI

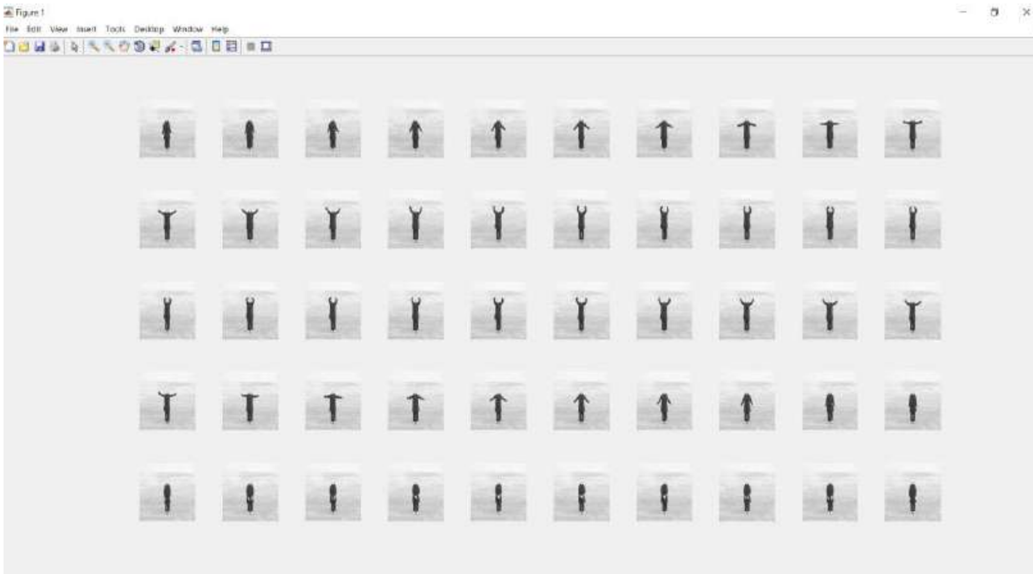

Fig. 8 The frames of a sample video of 'Hand Waving' of KTH dataset with Image AI

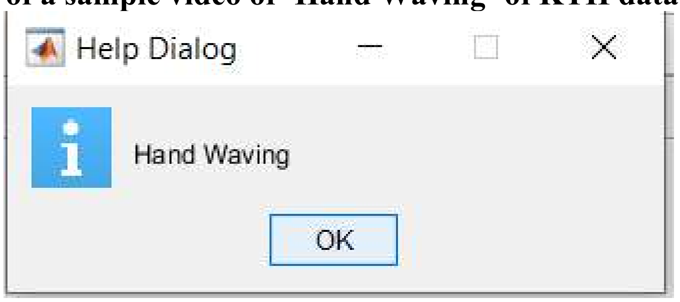

Fig. 9 The frames of a sample video of 'Hand Waving

Conclusion

The fundamental purpose of human action recognition is to look for continuing human activities in photos or video (a continuous sequence of frames). In this paper, we present an enhanced HAR model based on procedures such as deleting superfluous frames from movies, extracting SoIs, and using feature descriptors. The implementation of a KNN machine learning algorithm for the recognition of daily human actions was described in this study. This algorithm has a 97.46 percent testing accuracy and a 2.54 percent testing loss rate. High-speed processors are insufficient for processing deep learning models, which are tensors of extremely enormous size that are formed after preprocessing datasets. Because the datasets used in deep learning video processing are typically huge, advanced processing requires GPU computing with better performance.

References

[1] D. Thombre, J. Nirmal, and D. Lekha, "Human detection and tracking using image segmentation and Kalman filter," in 2009 International Conference on Intelligent Agent & Multi-Agent Systems, 2009, pp. 1-5: IEEE.

[2] M. Sharif, M. A. Khan, T. Akram, M. Y. Javed, T. Saba, and A. Rehman, "A framework of human detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropybased features selection," EURASIP Journal on Image and Video Processing, vol. 2017, no. 1, p. 89, 2017.

[3] J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu, "Automatic traffic surveillance system for vehicle tracking and classification," IEEE Transactions on Intelligent Transportation Systems, vol. 7, no. 2, pp. 175- 187, 2006.
[4] G. C. Dedes and K. C. Mouskos, "GPS/IMU/video/radar absolute/relative positioning communication/computation sensor platform for automotive safety applications," ed: Google Patents, 2014.
[5] G. Kapidis, R. Poppe, E. van Dam, L. P. Noldus, and R. C. Veltkamp, "Egocentric Hand Track and Object-based Human Action Recognition," arXiv preprint arXiv:1905.00742, 2019.
[6] A. Veenendaal, E. Daly, E. Jones, Z. Gang, S. Vartak, and R. S. Patwardhan, "Sensor tracked points and HMM based classifier for human action recognition," Computer Science and Emerging Research Journal, vol. 5, pp. 4-8, 2016.
[7] J. Cacace, A. Finzi, V. Lippiello, M. Furci, N. Mimmo, and L. Marconi, "A control architecture for multiple drones operated via multimodal interaction in search & rescue mission," in 2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 2016, pp. 233-239: IEEE.
[8] M. A. Goodrich and A. C. Schultz, "Human–robot interaction: a survey," Foundations and Trends® in Human–Computer Interaction, vol. 1, no. 3, pp. 203-275, 2008.
[9] Q. Ye, Z. Han, J. Jiao, and J. Liu, "Human detection in images via piecewise linear support vector machines," IEEE transactions on image processing, vol. 22, no. 2, pp. 778-789, 2012.
[10] S. Ali, A. Basharat, and M. Shah, "Chaotic invariants for human action recognition," in 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1-8: IEEE.
[11] V. Parameswaran and R. Chellappa, "View invariance for human action recognition," International Journal of Computer Vision, vol. 66, no. 1, pp. 83-101, 2006.
[12] A. Yilmaz and M. Shah, "Recognizing human actions in videos acquired by uncalibrated moving cameras," in Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 2005, vol. 1, pp. 150-157: IEEE.
[13] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 1, pp. 65-81, 2006.
[14] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8: IEEE.
[15] A. Agarwal and B. Triggs, "Recovering 3D human pose from monocular images," IEEE transactions on pattern analysis and machine intelligence, vol. 28, no. 1, pp. 44- 58, 2005.
[16] M. V. da Silva and A. N. J. A. S. C. Marana, "Human action recognition in videos based on spatiotemporal features and bag-of-poses," p. 106513, 2020.
[17] M. F. Aslan, A. Durdu, K. J. N. C. Sabanci, and Applications, "Human action recognition with bag of visual words using different machine learning methods and hyperparameter optimization," vol. 32, no. 12, pp. 8585-8597, 2020.
[18] I. M. Nasir, M. A. Khan, A. Armghan, and M. Y. Javed, "SCNN: A Secure Convolutional Neural Network using Blockchain," in 2020 2nd International Conference on Computer and Information Sciences (ICCIS), 2020, pp. 1- 5: IEEE.
[19] I. M. Nasir et al., "Pearson Correlation-Based Feature Selection for Document Classification Using Balanced Training," vol. 20, no. 23, p. 6793, 2020.

[20] I. M. Nasir, M. Rashid, J. H. Shah, M. Sharif, M. Y. H. Awan, and M. H. J. C. m. i. Alkinani, "An Optimized Approach for Breast Cancer Classification for Histopathological Images Based on Hybrid Feature Set," 2020.

[21] I.-M. Nasir et al., "Deep Learning-Based Classification of Fruit Diseases: An Application for Precision Agriculture," vol. 66, no. 2, pp. 1949-1962, 2021.

[22] I.-M. Nasir, M.-A. Khan, M. Alhaisoni, T. Saba, A. Rehman, and T. Iqbal, "A Hybrid Deep Learning Architecture for the Classification of Superhero Fashion Products: An Application for Medical-Tech Classification," p. {pages}.

[23] M.-A. Khan et al., "A Blockchain Based Framework for Stomach Abnormalities Recognition," vol. 67, no. 1, pp. 141-158, 2021.

[24] N. Yudistira and T. Kurita, "Correlation Net: Spatiotemporal multimodal deep learning for action recognition," Signal Processing: Image Communication, vol. 82, p. 115731, 2020.

[25] J. Wang, X. Peng, and Y. Qiao, "Cascade multi-head attention networks for action recognition," Computer Vision and Image Understanding, p. 102898, 2020.

[26] A. Sasithradevi and S. M. M. Roomi, "Video classification and retrieval through spatio-temporal Radon features," Pattern Recognition, vol. 99, p. 107099, 2020.

[27] C. Dai, X. Liu, and J. J. A. s. c. Lai, "Human action recognition using two-stream attention based LSTM networks," vol. 86, p. 105820, 2020.

[28] M. A. Khan, T. Akram, M. Sharif, M. Y. Javed, N. Muhammad, and M. Yasmin, "An implementation of optimized framework for action classification using multilayers neural network on selected fused features," Pattern Analysis and Applications, vol. 22, no. 4, pp. 1377-1397, 2019.

[29] J. C. Nunez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Velez, "Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition," Pattern Recognition, vol. 76, pp. 80-94, 2018.

[30] M. Ma, N. Marturi, Y. Li, A. Leonardis, and R. Stolkin, "Region-sequence based six-stream CNN features for general and fine-grained human action recognition in videos," Pattern Recognition, vol. 76, pp. 506-521, 2018.