Computers, Materials & Continua
DOI: 10.32604/cmc.2023.035894
Article

Tech Science Press

Human Verification over Activity Analysis via Deep Data Mining

Kumar Abhishek[1,*] and Sheikh Badar ud din Tahir[2]

[1]Senior Member, IEEE, Seattle, 98103, USA
[2]Department of Software Engineering, Capital University of Science and Technology (CUST), Islamabad, 44000, Pakistan
*Corresponding Author: Kumar Abhishek. Email: kr.abhishek@ieee.org

Received: 08 September 2022; Accepted: 08 December 2022

Abstract: Human verification and activity analysis (HVAA) are primarily employed to observe, track, and monitor human motion patterns using red-green-blue (RGB) images and videos. Interpreting human interaction using RGB images is one of the most complex machine learning tasks in recent times. Numerous models rely on various parameters, such as the detection rate, position, and direction of human body components in RGB images. This paper presents robust human activity analysis for event recognition via the extraction of contextual intelligence-based features. To use human interaction image sequences as input data, we first perform a few denoising steps. Then, human-to-human analyses are employed to deliver more precise results. This phase follows feature engineering techniques, including diverse feature selection. Next, we used the graph mining method for feature optimization and AdaBoost for classification. We tested our proposed HVAA model on two benchmark datasets. The testing of the proposed HVAA system exhibited a mean accuracy of 92.15% for the Sport Videos in the Wild (SVW) dataset. The second benchmark dataset, UT-interaction, had a mean accuracy of 92.83%. Therefore, these results demonstrated a better recognition rate and outperformed other novel techniques in body part tracking and event detection. The proposed HVAA system can be utilized in numerous real-world applications including, healthcare, surveillance, task monitoring, atomic actions, gesture and posture analysis.

Keywords: AdaBoost; classification; deep features mining; graph mining; human detection; human verification

1 Introduction

The modern age is characterized by innovation in a wide variety of fields, including information systems, machine learning, smart and intelligent systems, prediction and estimation-based frameworks, and automation. These fields provide opportunities for the sustainable development of advanced systems and intelligent tools for gathering data, from conventional camera systems to motion-based detectors. These intelligent technologies allow us to explore ideas in a variety of disciplines. One avenue of exploration is to discover and evaluate human verification technologies that may be used to improve living standards in communities around the world. However, this field still has limitations and problems, including noise removal, position prediction, human recognition, human activity analysis, motion prediction, feature extraction, data optimization, and classification of distinct activities. With the involvement of previous information and technological advancements, these difficulties must be addressed.

Recently, motion detectors and camera-based human movement recognition algorithms have been utilized in several scientific fields and intelligent applications [1–3]. Although many of these are part of intelligent devices such as those found in smart homes, [4] the internet of things, machine-learning-based numerical modelling [5], network security and encrypted communications, intelligent emergency frameworks, academic and monitoring solutions, e-learning strategies, smart transportation infrastructure, and intelligent medical frameworks, they are not smart frameworks themselves. In the academic system, administrators may watch the activity of students. However, in the cyber security field, these systems are used to detect the typical activity of humans and bots as well as find anomalies. In sports, they can be used to analyze player and crowd activity [6]. Using this approach, we can also analyze the human activity of patients, doctors, and visitors in the medical domain [7,8] or perform human recognition in home automation [9,10] or Internet of Things-based platforms.

In this work, we designed an effective approach for human identification and verification as well as human activity analysis in various indoor and outdoor settings. Primarily, we used indoor and outdoor video-based data as input to the proposed research approach. After pre-processing human shapes, we perform human verification and human activity analysis. For activity analysis, we must extract the context of intelligent features over the video-based dataset. To deal with the associated computational costs, we used a deep features mining approach via graph mining. Then, to analyze the human activity, we adopted a machine learning-based AdaBoost algorithm. We used two publicly-accessible datasets, U-T interaction and the Sports Videos in the Wild dataset. Fig. 1 shows the overall description and architecture of our study.

Using these two datasets, we achieved a significantly higher recognition rate than the other state-of-the-art techniques. The following is an overview of the key contributions and improvements in this study:

- We developed a comprehensive strategy for human verification and activity analysis using indoor and outdoor video-based data as well as various human involvement situations.
- We examined two human detection and verification algorithms to obtain more accurate, robust results; this is the primary consideration of several useful applications. The proposed technique helps us to get accurate information regarding human activity prediction.
- Context intelligent features are adopted for human verification and activity analysis. Furthermore, we used the deep features mining approach via a graph mining algorithm and activity prediction using an AdaBoost algorithm.
- The performance and effectiveness of the proposed system are illustrated through experimental observations over two publicly-accessible datasets. This shows that our research has significantly outperformed existing state-of-the-art methods.

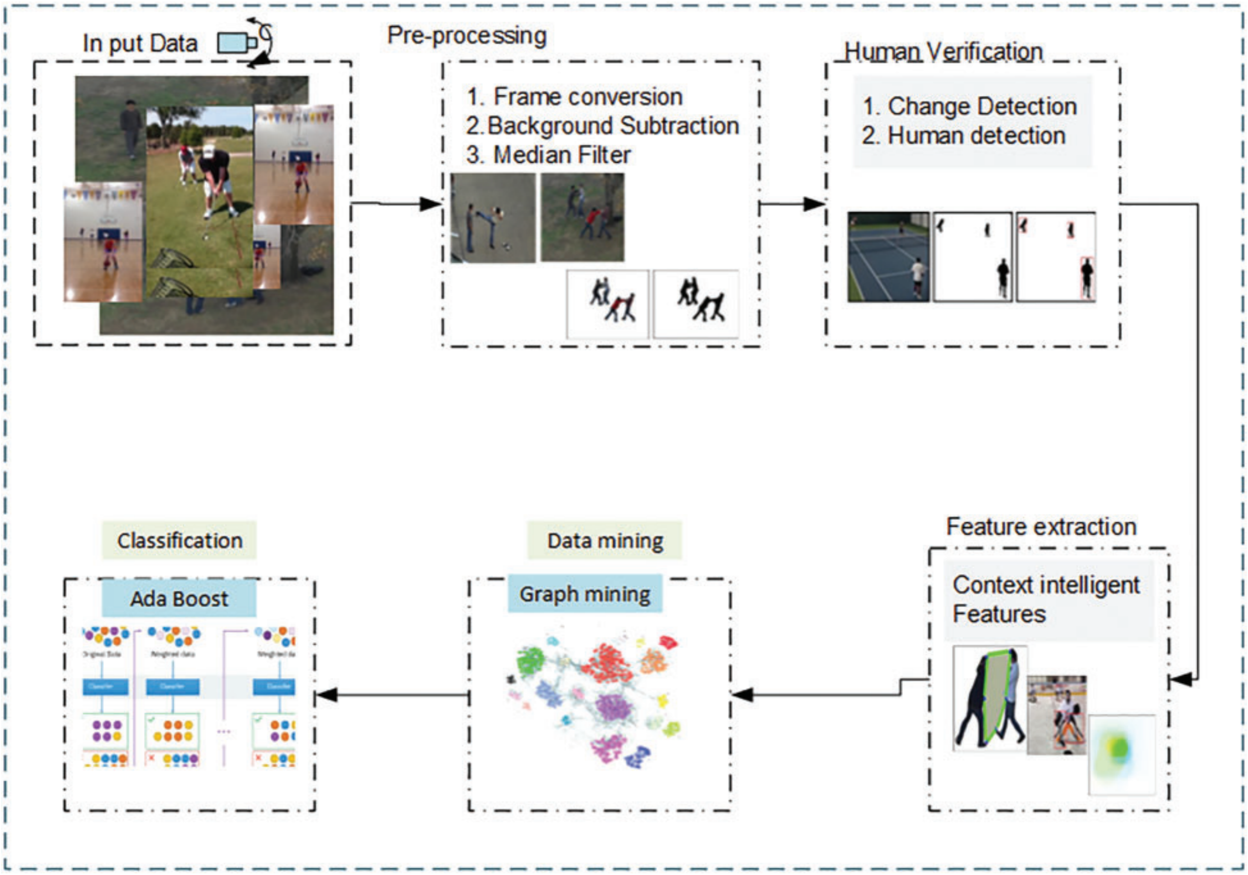


Figure 1: Detailed overview of the proposed architecture via graph-based deep features mining and AdaBoost classification

The remainder of our work is organized as follows: Section 2 discusses the related and previous work. Section 3 shows the flow design of the proposed approach, which consists of pre-processing, human detection, feature extraction, and deep features mining via graph mining. And classification using the AdaBoost classification algorithm. Section 4 describes the comprehensive analysis and evaluation of two state-of-the-art datasets, such as the Sports Video in the Wild and U-T interaction. After this. The detailed comparison existing system. Finally, Section 6 shows the paper's conclusion, limitations and future directions.

2 Objective 1: Background Research and Literature Review Study

Developments in cell phone cameras and live streams, as well as advancements in object indicator motion gadgets, allow for improved data farming and collection while numerous research institutions focus on extracting features and human action recognition studies [11].

Wang et al. [12] recently developed a novel human activity analysis (HAA) approach for analyzing labelled statistics. The proposed technique has modified the current convolution neural network HAA by comparing the retrieved feature selection method. Through valuing consistency, the attention-oriented network design optimizes vital information while setting aside highly redundant and contradictory data. This HAA strategy is characterized by deep convolutional long short-term memory (LSTM) and convolution neural networks.

The results illustrated an improved performance, though on inaccurately specific data. The methodology has aided the technique of data stream categorization and used the simplest statistics. In [13], researchers created a compact strategic plan predicated on efficient allocation, illumination changes, and obtained image feature statistics. The researchers successfully accomplished human activity recognition and analysis via the conventional optimization procedure, body part identification, and compressed coefficient dictionary learning technique. Zhou et al. [14] recently introduced a novel HAA predicated model on a Bayesian convolution network (BCN), which allows every system to access data using either low-power back propagation connections or traditional radio frequency (RF) connectivity. Convolution layers are responsible for extracting the features. An autonomous decoder-a typical deep net classification-was added to improve its accuracy. In addition, the Bayesian network classified the security risks using the enhanced deep learning (EDL) framework and an efficient offloading strategy. The results indicated that the data was susceptible to multiple forms of ambiguity, such as cognitive ambiguity, which is referred to as durability and noise.

In [15], researchers expanded the computational infrastructure to support volumetric structures. Informed by learning psychology, the research identifies foreground patches as "key components" of the framework and asserts that they include abundant and distinct spatial features. Newell et al. [16] created an architecture resembling an hourglass for activity detection and appended a supervisory output to its base. The single individual posture problem identifies a single human stance with a basic environment and minimal distraction. Proposed techniques for estimating the posture of a single individual have had response and validity above 93%. Meanwhile, the majority of images contain numerous humans, making the single-person pose estimate methodology ineffective. Chen et al. [17] created a cascaded pyramid system to estimate the human activity of numerous individuals using a regression model and modification. The leading multi-person activity estimation algorithm subdivides the HAA and verification problem into several HAA and verification problems, which are then analyzed by object tracking and single-person identification and verification. This method is straightforward and reliable. However, its efficiency depends on the outcomes of object identification. Additionally, multiple people standings increase the diffraction issues.

Einfalt et al. [18] developed methods for predicting activities in the movement of sports players using multiple processes that extract 20 sequential posture frameworks from data that contains videos and sequential images. Considering translation activity classifications, researchers developed a neural sequence architecture for exact action analysis and recognition. Rado et al. [19] built a focused attentiveness (LSTM) framework that extracts CNN-based attributes and chronological positioning from challenging video sequences. To identify humans in images and video-based datasets, the YOLO v3 technique was formulated, whereas an LSTM-based technique was applied to identify anomalies. This research adopts supervised learning, convolutional neural network (CNN) techniques, or an insufficient number of attributes in multimedia databases to execute these methodologies. Franklin et al. [20] designed a complete deep learning system for classifying anomalous and routine activities. They utilized reduction, clustering, and graph-based methodologies to achieve relevant results. Through the deep learning method, the authors identified both normal and abnormal activity duration parameters. Additionally, Mishra et al. [21] propose a fractional derivative S-transform oriented feature-extraction and linear discriminant analysis (LDA) oriented feature reduction procedure. In addition, researchers have applied the AdaBoost method with random forests to the adequate detection of human activity.

Most of the proposed frameworks utilize supervised learning; fragmentation also plays a crucial part in the classification results. However, these procedures required that the humans and cells remain effectively segregated from input data.

Ghadi et al. [22] established a method for generating video characterization by combining a 3D CNN algorithm and LSTM-based decoder. They integrate the focus on visuals by establishing the likelihood function over the images, which is used to generate the actual words. Generally, studying the structure of a deep neural network-sometimes called a black box is challenging. Therefore, video feature descriptors acquire the model's focal point to improve the number of possible readings. In [23], the researcher described integrating movement energy projection and gait energy mapping to describe the action of the human body and, afterward, correlating the temporal pattern with the reference sequence. This technique effectively handles the given input data accurately and effectively, which has a negligible impact on the HAA. In another study [24], the author proposed an active learner incorporating Local Directional Pattern (LDP) as the representation to give the programmer more control over the feature extraction model. Additionally, the LDP framework was implemented for both simulated and actual active learning, achieving comparable performance.

Many of the research studies did not follow the pre-processing phase, which causes time complexity and resource requirements. In addition, numerous research studies incorporate traditional methods for human detection and recognition. Furthermore, they adopt a single technique to achieve this goal, which is less optimized. Moreover, researchers avoid extracting robust and multiple features and use a classification approach without data optimization. Due to these reasons, we face various issues such as less accuracy of the system, higher error rate, data normalization, optimization problems, and time-saving with resource utilization issues. To address aforementioned concerns, we created a robust framework to identify the human and analyze their activity in a human life log.

3 Objective 2: Proposed Methodology with Novelty Highlights

In this part, we present a complete explanation of the suggested system with detailed methods as well as results.

3.1 Methods

Initially, we transformed the video data into frames. Then, we reduced the converted frames to a set size, reduced distortion, and boosted image clarity. The next step was to detect the human from various structures and extract the following features: moveable body points, shape distance features, moveable body parts, and angular cosine features. After this, we needed to optimize the data for more accurate results. To achieve this, we applied graph mining. Finally, we used AdaBoost for classification and activity analysis.

3.2 Data Pre-Processing

Before human verification and identification, we utilized several pre-processing methods to reduce computational expense and time. This includes the preliminary conversion of video sequences to image data. These images have a constant size of 450 × 350 pixels. The images are then denoised via the median filtration process. Median filtering is performed to recognize deformed pixels in images and replace them with the median index. We used a 5 × 5 grid to reduce noise. The mathematical representation of the median filter is formulated in Eqs. (1)–(3):

$$Medf\ \ (I) = Medf\{I_m\} \tag{1}$$

$$= \frac{I_m\,(n+1)}{2}; n\ is\ odd \tag{2}$$

$$= \frac{1}{2}\left[I_m\left(\frac{n}{2}\right) + I_m\left(\frac{n}{2}\right) + 1\right], \tag{3}$$

where I1, I2, I3, . . . , In is the order of the adjacent pixels. All available pixels of the given images must be organized in order. Subsequently, the categorization of the pixels and the arrangement of the selected pixels will be $I_{m1} < I_{m2} < I_{m3} < I_{mn}$ where n is generally abnormal. Fig. 2 shows the results of noise reduction and data preprocessing.

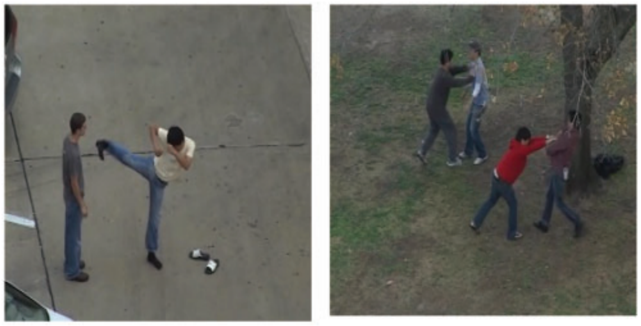


Figure 2: Example images after data preprocessing

3.3 Background Subtraction

For background removal, we used an improved joining methodology wherein we applied a Markov random field on color parameters and region fusion techniques. Next, we performed change recognition in-frame sequence using a dynamic threshold-based method followed by spatial-temporal variance to achieve more precise results. Fig. 3 depicted the results of background subtraction over the U-T interaction data set.

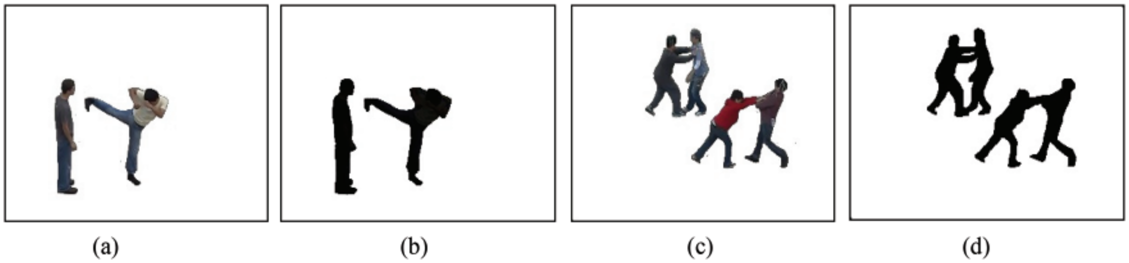


Figure 3: Background subtraction results in (a, c) background subtraction result and (b, d) binary conversion

3.4 Human Detection and Verification

In this section, we discuss the optimization of the identification of human silhouettes by combining change recognition, Markov random field, and spatial-temporal variance approaches with a dynamic thresholding strategy. Eq. (4) presents the equation we used for human head tracing,

$$T_H^w \leftarrow T_H^{w-1} + \Delta T_H^{w-1}, \tag{4}$$

where T_H^w characterizes a human head position in numerous specified input frames, w, which is calculated via spatial-temporal variance. For human recognition and verification, Eq. (5) demonstrates the following mathematical formulation:

$$T_{FH}^w = (T_H^w \leftarrow T_H^{w-1} + \Delta T_H^{w-1}) + T_{End}^w, \tag{5}$$

where T_{FH}^w characterizes a human position in numerous specified input frames, w, and T_{End}^w indicates the bounding container dimension for the human identification and verification Algorithm 1, which describes the human detection technique in detail. Fig. 4 shows the results of human detection and verification.

Algorithm 1: Human Detection and Verification

```
Input: ES: Extracted_Silhouettes of human
Output: Human detection and verification
/∗ human outer shape in input∗/, /∗ WR is for white region∗/, /∗ HS is human silhouette∗/
/∗ ShF is for shape feature∗/
Procedure: Repeat
For k = 1 to I do
        For k = 1 to I do
search (WR)
        End
End
If WR1 > WR
        WR = WR1
End
Until major entity figure explored
Step 2: /∗ Compare both WR ∗/
For all pixels in both WR
    If WR_(pixel_data_frame 1) = WR_(data_frame 2)
        WR_(pixel_data_frame 3) = WR_(pixel_data_frame 1)
    End
    If WR is inadequate for all inputs
        If pixel information is equal to ShF
            HS = WR_(pixel )
        End
    End
End
```

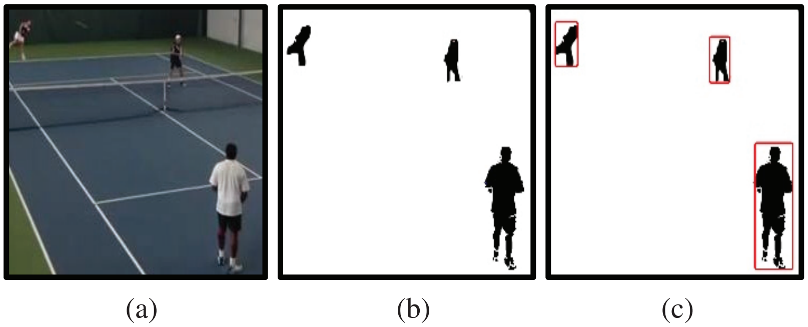


Figure 4: Human detection results in example images (a) original RGB image, (b) background subtraction result, and (c) human detection and verification

3.5 Feature Extraction

In this phase, we performed the extraction of features by extracting the moveable body parts, shape distance features, movement flow, and angular cosine features. Algorithm 2 shows the overview of feature extraction.

Features I: Shape distance features

In shape distance features, we extract the six points' index values and find the distance of all covered points over two human figures. These points are driven by the adopted approach according to the size and distance between the two humans. Eq. (6) presents the formulation of shape distance features,

$$Sdf = \binom{6}{2} * s * a, \tag{6}$$

where Sdf is the shape distance feature vector, 6/2 is a constant, s is the adjacent side of a hexagon and a represents the apothem distance. Fig. 5 shows the resulting shape distance features.

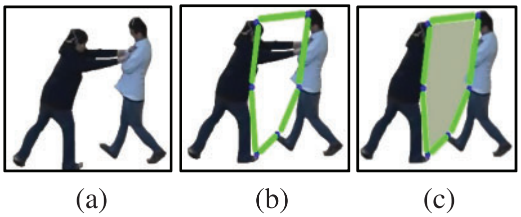


Figure 5: The results of shape distance features: (a) extracted background subtraction, (b) six points over humans, and (c) region of shape distance

Features II: Moveable body parts features

This technique targets the specific movable body components of humanoid shapes. Whenever human action is first identified, a mask is drawn around the movable section, and its pixel position is determined. Finally, we collect the images' top 35 values and translate them into verticals (see Fig. 6).

Features III: Angular cosine features

In angular cosine features, we map six points over the human body and join them as a mesh. With the help of trigonometric function, we find the angle of these extracted points and insert the resulting angular cosine feature's vector. Eq. (7) shows the mathematical formulation of the angular cosine features:

$$cos\,(x+y) = cos\,(x)\ *\ cos\,(y) - sin\,(x)\ *\ sin\,(y)\,, \tag{7}$$

where $cos\,(x+y)$ is the angle value and $x,\ y$ are point A and point B, respectively. Fig. 7 shows the detailed results of the angular cosine features.

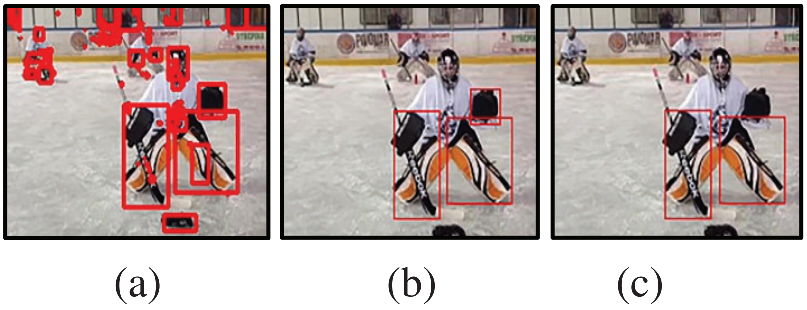


Figure 6: The results of moveable body parts features (a) general movement detection, (b) points to human and human-linked objects, and (c) moveable human objects

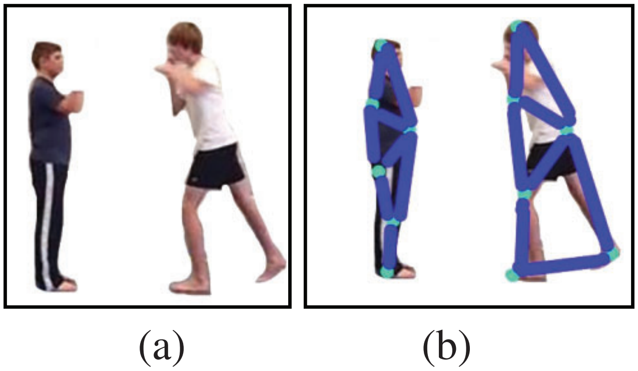


Figure 7: Results of angular cosine features: (a) background-subtracted image and (b) angular cosine features at every point

Features IV: Movement flow features

In movement flow features, we applied a color- and segmentation-based model over the given data frames. After this, we recognize the movement flow and mark it with various colors. Finally, we can get these index values and map them in vectors for future calculations and estimations.

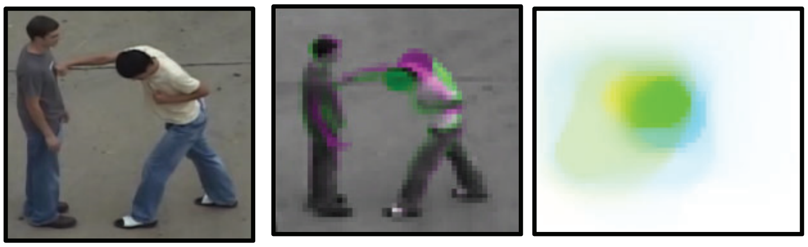
The movement flow features are formulated as

$$M_f = \sum\nolimits_0^n iv\,(pi)\,, \tag{8}$$

where M_f denotes the movement flow features vector, iv is the index RGB values, and pi is the given frame. Fig. 8 shows the results of movement flow features.

Algorithm 2: Feature Calculation

```
Input: Input_data
Output: Extracted Feature vectors ($f_{1,} f_{2,} f_3 \ldots f_n$)
Extracted_features ← []
F_Data ← Get_F_Data()
F_Data_size ← Get_F_Data_size()
Procedure HAA (Video, Images)
FeaturesVector ← []
Denoise_F_Data  ← Pre_processing (Win,Median)
Sampled_F_Data (DenoiseData)
While exit void state do
[$MBP$, $SDF$, $MFF$, $ACF$] ← ExtractlFeatures (sample data)
ExtractedFeaturesVector ← [$MBP$, $SDF$, $MFF$, $ACF$]
Return Context_intelligent_features_Vector
```



(a) (b) (c)

Figure 8: The results of movement flow feature: (a) original RGB image, (b) movement flow marked over the human body, and (c) movement values in various colors

3.6 Deep Feature Mining: Graph Mining

As features are retrieved from the entire dataset, the subsequent phase decreases the input indexes, which minimizes operational costs and enhances accuracy. To feature content that is also exposed to quantitative frameworks and indicators, scholars may have a high prediction result of retrieval by utilizing the graph mining methodology [25]. Graph mining combines methods and equipment for data processing, anticipating data models, and constructing an organized and realistic graph for pattern recognition. Algorithm 3 presents the whole functioning description of graph mining.

Algorithm 3: Data Mining via the Graph Mining Approach

```
Input: All Features (Af)
Output: Mined_data ($Mdt$)
All_feature ← []
for i = 1: k do
Read_Data: Q → (Af)
Tree_Creation: TC_tree (Q → 0)
```

(Continued)

Algorithm 3: Continued

```
Read_Data: to find min R(min) and max R(Max)
Find_next_node: R(Af → next_node)
Find Mutual_node: apprise_the_list
Mine_the_date: min (Tree, apprise)
Restrictive_TC_tree:Produce_the_tree (mining)
end
return Optimized Data {OD}
```

3.7 *Classification: AdaBoost*

AdaBoost is one of the most frequently used techniques for classification; it builds a robust classifier by using a joint distribution of several component models. AdaBoost is used to determine low-quality participants to construct a group statistical method. During the training phase, the member classifiers are selected to achieve the lowest possible margin of error for each category.

In the second phase of recurrence, AdaBoost presents a technique that is not only straightforward to use, but also adequate for the generation of ensemble methods. This is done using a recurrent phase adjustment over the entire training collection, one of its hybrid properties [26]. Eq. (9) shows the formulation of the training phase of the AdaBoost algorithm.

$$M_{q(x)}\sum_{n=1}^{q} a_n\left(x\right), \tag{9}$$

where a_n is an enhanced learner that generates a feature x as a contribution and computes the worth to recognize an entity class. At each recurrence of the training technique, a weight, $we_{i,n}$, is assigned to each segment in the input set that is identical to the obtainable inaccuracy, $Er\left(B_{n-1}\left(a_i\right)\right)$, on that segment. Here, $B_{n-1}(a_i)$ is represented as a boosted classifier, which recognizes the fragile learner. Fig. 9 shows the AdaBoost model diagram.

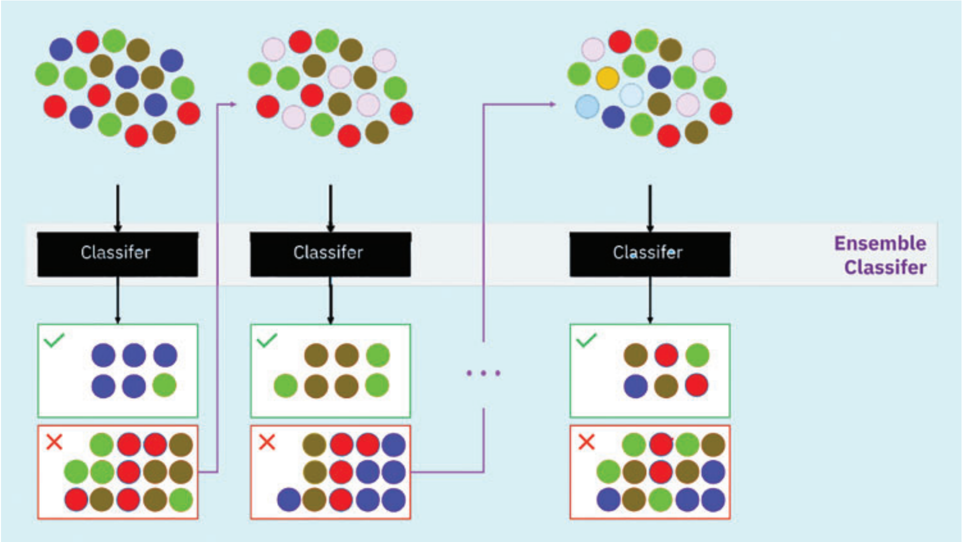


Figure 9: Detailed model of the AdaBoost classification algorithm

4 Objective 3: Experimental Results

The leave-one-subject-out (LOSO) cross-validation technique has been employed to test the performance of the HVAA system over two openly accessible datasets, including the Sports Videos in the Wild (SVW) dataset and the UT-interaction dataset. The LOSO technique is a modified cross-validation method that involves single-subject data for each fold.

4.1 Datasets Information

The benchmark datasets include diverse sports activities and sophisticated human-human interaction scenes. In the SVW dataset [27], most of the videos were captured with the Coach's Eye mobile app, an innovative sports training program developed by TechSmith specifically for smartphones. There are 19 activity categories for 19 various activities, including *archery, baseball, basketball, BMX, bowling, boxing, cheerleading, football, golf, high jump, hockey, hurdling, javelin, long jump, pole vault, rowing, shotput, skating, tennis, volleyball, and weightlifting*; all images were acquired at a resolution of 720 × 480 and at 30 frames per second. Fig. 10 depicts a sampling of photos from the SVW collection.

Figure 10: Example images from various classes of the SVW dataset

The second benchmark UT-interaction dataset [28] contains recordings of six classes of periodically conducted human-human interactions: shaking hands, pointing, hugging, driving, and striking. We accessed a preview of twenty one-minute-long video feeds. The expanded video data include at least one additional execution per contact, resulting in an average of eight human encounters per film. A large number of participants engage in the videos, which include more than 15 different outfits. The entire video was taken at a resolution of 720 × 480 at 30 frames per second. There are six distinct interaction classes: handshake, embrace, kick, point, punch and push. Fig. 11 illustrates a sampling of the photos from the UT-interaction dataset.

4.2 Experimental Results and Analysis

In the experiment of the proposed HVAA system, we used MATLAB (R2021a) for all simulations and estimations. We also used an Intel (R) Core (TM) i7-8665U @ 1.90 GHz CPU with 64 bit Windows 11 in the testing device. The device had 16 GB of RAM. The new verdict on the SVW and UT-interaction datasets along with experimental outcomes is analyzed in the results section.

Figure 11: Example images from various classes of the U-T interaction dataset

Experimental Setup and Evaluation

We undertook two tests to evaluate the performance of the proposed HVAA system across two benchmark datasets. Tables 1 and 2 display the real subject count and human verification average accuracy based on the variation of the frame data. Table 1 had five columns, the first of which represents the series of specified frames, the next represents the real track, the third contains successful tracking, the fourth column shows the number of failures, and the fifth indicates the accuracy of the SVW and UT-interaction datasets.

Table 1: Actual human detection and verification accuracy over the SVW dataset

Frames	Real track	Successful	Failure	Accuracy
8	3	3	0	100
16	3	3	0	100
24	5	5	0	100
32	5	4	1	80
40	7	6	1	85.71
48	9	8	1	88.88
56	9	8	1	88.88

Table 2: Actual human detection and verification accuracy over the UT-Interaction dataset

Frames	Real track	Successful	Failure	Accuracy
8	11	11	0	100
16	11	9	2	81.81
24	11	10	1	90.90
32	14	13	1	92.85
40	14	13	1	92.85
48	17	15	2	88.23
56	17	16	1	94.11

The next stage of research was to determine the typical and abnormal events of the proposed HVAA system with the assistance of the AdaBoost model algorithm. Fig. 12 presents the confusion matrix for the SVW dataset with a recognition rate of 92.15%. Fig. 13 shows the confusion matrix for the UT-interaction dataset, which has an average accuracy of 92.83%.

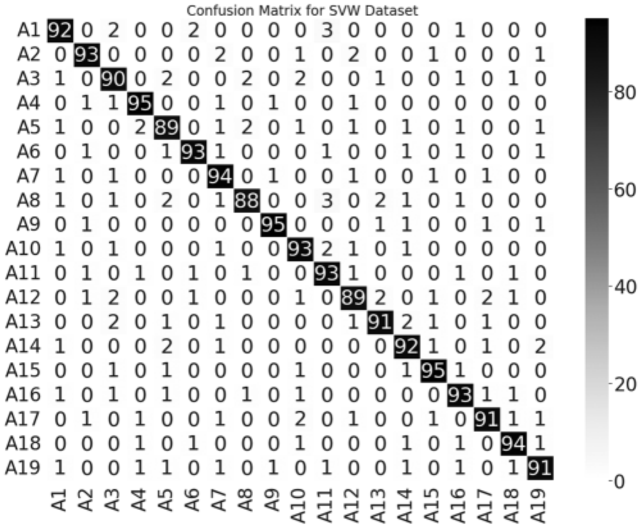


Figure 12: Confusion matrix of 19 different genre sports activities on the SVW dataset (Note: A1 = archery, A2 = baseball, A3 = basketball, A4 = bmx, A5 = bowling, A6 = cheerleading, A7 = football, A8 = golf, A9 = highjump, A10 = hockey, A11 = hurdling, A12 = javelin, A13 = longjump, A14 = polevault, A15 = rowing, A16 = shotput, A17 = skating, A18 = tennis, A19 = volleyball)

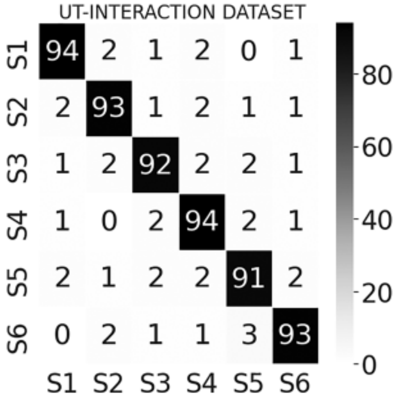


Figure 13: Confusion matrix of 6 distinct interaction activities on the UT-INTERACTION dataset (Note: S1 = hand_shaking, S2 = hugging, S3 = kicking, S4 = pointing, S5 = punching, S6 = pushing)

Objective 4: Performance comparison.

Experiment II: Comparison with other Algorithms

After achieving significant mean recognition results for our proposed HVAA system, we compared it with novel classification techniques. Table 3 reveals that our performance [29] on the benchmark datasets is significantly higher than the results from previous techniques. For example, the framework of Markov random field is adopted by Park et al. [30], which combines pixels into interconnected blobs and tracks inter-blob correlations. On the other hand, the conventional neural network introduced by Li et al. [31] estimated the human body pose. Additionally, Chen et al. [32] employed morphological segmentation of the top color along with methodical thresholding. Rodriguez et al. [33] also developed a new approach for predicting future body movement. Additionally, they incorporated logical explanations and focused failure mechanisms to support a regenerative system that forecasts definite future human motion. The evaluation of complex event detection and classification with state-of-the-art techniques is presented in Table 3. In Tables 4 and 5, we evaluated the performance of the proposed HVAA system by comparing it with two other state-of-the-art methods, namely, Maximum Entropy Markov Model (MEMM) and Genetic Algorithm (GA) classifiers. We compared their precision, their recall, and the F1 scores of all classes used in the two benchmark datasets, SVW and UT-Interaction.

Table 3: Event classification comparison of recognition rate of the HVAA proposed method with other state-of-the-art methods over UT and SVW datasets

Frameworks	UT (%)	Frameworks	SVW (%)
Rodriguez et al. [33]	71.80	Zhu et al. [34]	83.10
Xing et al. [35]	85.67	Rachmadi et al. [36]	82.30
Chattopadhyay et al. [37]	89.25	Sun et al. [38]	74.20
Proposed HVAA	**92.83**		**92.15**

Table 4: Measurements of evaluation metrics for the proposed HVAA system over the SVW dataset

SVW	AdaBoost			Maximum entropy markov model			Genetic algorithm		
Activities	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
A1	0.920	0.920	0.920	0.890	0.880	0.885	0.875	0.870	0.872
A2	0.939	0.930	0.934	0.912	0.910	0.911	0.901	0.900	0.900
A3	0.882	0.900	0.891	0.810	0.840	0.825	0.864	0.870	0.867
A4	0.940	0.950	0.945	0.921	0.900	0.910	0.911	0.920	0.915
A5	0.936	0.890	0.912	0.897	0.910	0.903	0.878	0.890	0.884
A6	0.948	0.930	0.939	0.865	0.880	0.872	0.858	0.860	0.859
A7	0.895	0.940	0.917	0.858	0.870	0.864	0.876	0.880	0.878
A8	0.936	0.880	0.907	0.898	0.910	0.904	0.882	0.890	0.886
A9	0.969	0.950	0.959	0.924	0.940	0.932	0.897	0.900	0.898
A10	0.902	0.930	0.916	0.875	0.930	0.902	0.869	0.870	0.869
A11	0.902	0.930	0.916	0.873	0.860	0.866	0.868	0.920	0.893
A12	0.908	0.890	0.899	0.896	0.940	0.917	0.875	0.860	0.867

(Continued)

Table 4: Continued

SVW	AdaBoost			Maximum entropy markov model			Genetic algorithm		
Activities	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
A13	0.938	0.910	0.924	0.912	0.930	0.921	0.904	0.860	0.881
A14	0.901	0.920	0.910	0.883	0.840	0.861	0.879	0.890	0.884
A15	0.940	0.950	0.945	0.923	0.910	0.916	0.875	0.910	0.892
A16	0.911	0.930	0.920	0.897	0.880	0.888	0.921	0.930	0.925
A17	0.928	0.910	0.919	0.895	0.920	0.907	0.893	0.910	0.901
A18	0.94	0.940	0.940	0.901	0.900	0.900	0.886	0.870	0.878
A19	0.919	0.910	0.914	0.890	0.880	0.885	0.897	0.890	0.893

Table 5: Measurements of evaluation metrics of the proposed HVAA system over the UT-Interaction dataset

SVW	AdaBoost			Maximum entropy markov model			Genetic algorithm		
Activities	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
S1	0.940	0.940	0.940	0.913	0.910	0.911	0.909	0.91	0.909
S2	0.930	0.930	0.930	0.904	0.900	0.902	0.906	0.900	0.903
S3	0.929	0.920	0.924	0.917	0.920	0.918	0.901	0.910	0.905
S4	0.912	0.940	0.926	0.909	0.900	0.904	0.912	0.920	0.916
S5	0.919	0.910	0.914	0.915	0.920	0.917	0.901	0.930	0.915
S6	0.939	0.930	0.934	0.925	0.930	0.927	0.917	0.920	0.918

In Tables 4 and 5, we compared the SVW and UT-Interaction datasets with the Maximum Entropy Markov Model (MEMM) and the Genetic Algorithm (GA). These results show that AdaBoost achieves better classification scores (precision, recall, and F-measure) when employed for predicting and classifying extraneous human behavior.

Due to the complex nature of these benchmark datasets, this study has one drawback, namely, occlusion. This problem impacts human tracking and verification as well as the feature engineering process. This is the main factor that caused the mean recognition to drop.

5 Discussion

Our HVAA system is designed to predict the extraneous interactions of human activities in various indoor and outdoor environments using graph features-based mining and AdaBoost classification. This study focuses on denoising, human interaction and verification, multi-subject analyses, feature engineering, feature selection, and behavior analysis. Initially, we conducted the pre-processing phase to lower computational overhead, as some of the data involved both human and non-human random objects simultaneously. To mitigate this issue, human-related verification and robust multi-person were conducted. For multiclass classification and estimation, feature engineering is a significant step. We introduced robust, contextually intelligent features as well as deep mining features. Additionally, a graph mining strategy was applied for feature optimization. Finally, AdaBoost was employed for predicting and classifying extraneous human behavior.

Due to the complex nature of these benchmark datasets, this study has one drawback, namely, occlusion. This problem impacts human tracking and verification as well as the feature engineering process. This is the main factor that caused the mean recognition to drop.

6 Analysis

In this section, we critically analyze our proposed methodology. Initially, we present a robust approach to overcome the research gaps, such as the human detection methods. Then, we perform various algorithms to detect humans and optimize them to attain accurate results. Next, we provide multiple feature extraction approaches for abstracting valuable data. After feature representation, we optimize them through optimization algorithms. For classification, we utilized Adaboost in order to get more accurate results than existing methods.

7 Conclusion and Future Insight

Our proposed work presented a step forward in the system to predict human behavior and determine both normal and abnormal events in an indoor-outdoor environment. First, we used two benchmark datasets as the input stream via numerous preprocessing techniques. These datasets involved sports and event-sourced information. Second, we denoised the sequence of images and dimensions, tracking the human and non-human objects. Following that, we performed feature engineering to extract diverse features. Next, we employed the graph-mining strategy to reduce the computational overhead and improve the recognition rate. Finally, the AdaBoost model was incorporated to predict the activities and locomotion patterns of numerous subjects. This study also compares the performance of our HVAA proposed system with that of other state-of-the-art methods. The experimental results have shown significant performance improvement over two benchmark datasets when compared with other state-of-the-art techniques.

In future work, we will integrate more complex tasks from various contexts, including medical centers, workplaces, IoT based system, Security and surveillance based system and smart homes. Additionally, we will fuse more feature engineering techniques from different domains in order to recognize complex motion patterns in multiple contexts along with human 3D modeling and 3D image reconstruction.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. A. ur Rehman, H. Raza and I. Akhter, “Security enhancement of hill cipher by using non-square matrix approach,” in *Proc. Conf. on Knowledge and Innovation in Engineering, Science and Technology*, Berlin, Germany, pp. 1–7, 2018.

[2] S. B. ud din Tahir, “A triaxial inertial devices for stochastic life-log monitoring via augmented-signal and a hierarchical recognizer,” Doctoral Dissertation, M.S. Thesis, Dept. Computer science, Air University, Islamabad, Pakistan, 2020.

[3] Y. Ghadi, I. Akhter, M. Alarfaj, A. Jalal and K. Kim, "Syntactic model-based human body 3D reconstruction and event classification via association based features mining and deep learning," *PeerJ Compututer Science*, vol. 7, pp. e764, 2021.

[4] D. Bhargavi, E. P. Coyotl and S. Gholami, "Knock, knock. Who's there?–Identifying football player jersey numbers with synthetic data," arXiv preprint arXiv:2203.00734, 2022.

[5] L. Kratz and K. Nishino, "Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 987–1002, 2011.

[6] S. Gholami and M. Noori, "You don't need labeled data for open-book question answering," *Applied Sciences*, vol. 12, no. 1, pp. 111, 2022.

[7] A. Ahmed, A. Jalal and A. A. Rafique, "Salient segmentation based object detection and recognition using hybrid genetic transform," in *2019 Int. Conf. on Applied and Engineering Mathematics (ICAEM)*, Islamabad, Pakistan, pp. 203–208, 2019.

[8] S. B. ud din Tahir, A. B. Dogar, R. Fatima, A. Yasin, M. Shafiq *et al.,* "Stochastic recognition of human physical activities via augmented feature descriptors and random forest model," *Sensors*, vol. 22, no. 17, pp. 6632, 2022.

[9] A. Ahmed, A. Jalal and K. Kim, "RGB-D images for object segmentation, localization and recognition in indoor scenes using feature descriptor and hough voting," in *2020 17th Int. Bhurban Conf. on Applied Sciences and Technology (IBCAST)*, Islamabad, Pakistan, pp. 290–295, 2020.

[10] J. Wang and Z. Xu, "Spatio-temporal texture modelling for real-time crowd anomaly detection," *Computer Vision and Image Understanding*, vol. 144, pp. 177–187, 2016.

[11] I. Akhter, A. Jalal and K. Kim, "Adaptive pose estimation for gait event detection using context-aware model and hierarchical optimization," *Journal of Electrical Engineering & Technology*, vol. 9, pp. 1–9, 2021.

[12] K. Wang, J. He and L. Zhang, "Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors," *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7598–7604, 2019.

[13] A. Li, Z. Miao, Y. Cen, X. -P. Zhang, L. Zhang *et al.,* "Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning," *Pattern Recognition*, vol. 108, pp. 107355, 2020.

[14] Z. Zhou, H. Yu and H. Shi, "Human activity recognition based on improved Bayesian convolution network to analyze health care data using wearable IoT device," *IEEE Access*, vol. 8, pp. 86411–86418, 2020.

[15] N. T. Thành, P. T. Công and L. V. Hung, "An evaluation of pose estimation in video of traditional martial arts presentation," *Journal on Information Technologies & Communications*, vol. 2019, no. 2, pp. 114–126, 2019.

[16] A. Newell, K. Yang and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. of the 14th European Conf. ECCV*, Amsterdam, The Netherlands, pp. 483–499, 2016.

[17] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu *et al.,* "Cascaded pyramid network for multi-person pose estimation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 7103–7112, 2018.

[18] M. Einfalt, C. Dampeyrou, D. Zecha and R. Lienhart, "Frame-level event detection in athletics videos with pose-based convolutional sequence networks," in *Proc. of the 2nd Int. Workshop on Multimedia Content Analysis in Sports-MMSports '19*, New York, NY, USA, pp. 42–50, 2019.

[19] D. Rado, A. Sankaran, J. Plasek, D. Nuckley and D. F. Keefe, "A Real-time physical therapy visualization strategy to improve unsupervised patient rehabilitation," in *Proc. of the IEEE Transactions on Visualization and Computer Graphics*, Atlantic City, NJ, USA, pp. 1–2, 2012.

[20] R. J. Franklin, Mohana and V. Dabbagol, "Anomaly detection in videos for video surveillance applications using neural networks," in *Proc. of the 2020 Fourth Int. Conf. on Inventive Systems and Control (ICISC)*, Coimbatore, India, pp. 632–637, 2020.

[21] S. Mishra, B. Majhi and P. K. Sa, "Glrlm-based feature extraction for acute lymphoblastic leukemia (all) detection," in *Recent Findings in Intelligent Computing Techniques*, Berlin/Heidelberg, Germany: Springer, pp. 399–40, 2018.

[22] Y. Y. Ghadi, I. Akhter, S. A. Alsuhibany, T. al Shloul, A. Jalal *et al.,* "Multiple events detection using context-intelligence features," *Intelligent Automation & Soft Computing*, vol. 34, no. 3, pp. 1455–1471, 2022.

[23] B. R. Murlidhar, R. K. Sinha, E. T. Mohamad, R. Sonkar and M. Khorami, "The effects of particle swarm optimisation and genetic algorithm on ANN results in predicting pile bearing capacity," *International Journal of Hydromechatronics*, vol. 3, no. 1, pp. 69–87, 2020.

[24] Q. Xiao and R. Song, "Human motion retrieval based on statistical learning and Bayesian fusion," *PLoS One*, vol. 11, no. 10, pp. e0164610, 2016.

[25] D. Chakrabarti and C. Faloutsos, "Graph mining: Laws, generators, and algorithms," *ACM Comput. Surv.*, vol. 38, no. 1, pp. 2–es, 2006.

[26] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999.

[27] S. M. Safdarnejad, X. Liu, L. Udpa, B. Andrus, J. Wood *et al.,* "Sports videos in the wild (SVW): A video dataset for sports analysis," in *Proc. of the 2015 11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana, Slovenia, pp. 1–7, 2015.

[28] M. S. Ryoo, C. -C. Chen, J. K. Aggarwal and A. Roy-Chowdhury, "An overview of contest on semantic description of human activities (SDHA) 2010," in *Recognizing Patterns in Signals, Speech, Images and Videos*, Berlin/Heidelberg, Germany: Springer, pp. 270–285, 2010.

[29] I. Akhter, A. Jalal and K. Kim, "Pose estimation and detection for event recognition using sense-aware features and adaboost classifier," in *Proc of. Conf. on Applied Sciences and Technologies (IBCAST)*, Islamabad, Pakistan, pp. 500–505, 2021.

[30] S. Park and J. K. Aggarwal, "Segmentation and tracking of interacting human body parts under occlusion and shadowing," in *Proc. of IEEE Workshop on Motion and Video Computing*, Orlando, FL, pp. 105–111, 2002.

[31] S. Li and A. B. Chan, "3D human pose estimation from monocular images with deep convolutional neural network," in *Proc. of the Asian Conf. on Computer Vision*, Singapore, pp. 332–347, 2014.

[32] H. W. Chen and M. McGurr, "Moving human full body and body parts detection, tracking, and applications on human activity estimation, walking pattern and face recognition," *Automatic Target Recognition XXVI*, vol. 984, pp. 213–246, 2016.

[33] C. Rodriguez, B. Fernando and H. Li, "Action anticipation by predicting future dynamic images," in *Proc. of the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 1–16, 2018.

[34] Y. Zhu, K. Zhou, M. Wang, Y. Zhao and Z. Zhao, "A comprehensive solution for detecting events in complex surveillance videos," *Multimedia Tools and Applications*, vol. 78, pp. 817–838, 2019.

[35] D. Xing, X. Wang and H. Lu, "Action recognition using hybrid feature descriptor and VLAD video encoding," in *Asian Conf. on Computer Vision*, Berlin/Heidelberg, Germany, Springer International Publishing, pp. 99–112, 2014.

[36] R. F. Rachmadi, K. Uchimura and G. Koutaki, "Combined convolutional neural network for event recognition," in *Proc. of the Korea-Japan Joint Workshop on Frontiers of Computer Vision*, Takayama, Japan, pp. 85–90, 2016.

[37] C. Chattopadhyay and S. Das, "Supervised framework for automatic recognition and retrieval of interaction: A framework for classification and retrieving videos with similar human interactions," *IET Computer Vision*, vol. 10, pp. 220–227, 2016.

[38] S. Sun, Z. Kuang, L. Sheng, W. Ouyang and W. Zhang, "Optical flow guided feature: A fast and robust motion representation for video action recognition," in *Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 1390–1399, 2018.