Abnormal Human Activity Recognition and Classification Using Knowledge Graphs

Sunitha Sabbu*[ID], Vithya Ganesan[ID]

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram A.P 522502, India

Corresponding Author Email: sunindra111@gmail.com

ABSTRACT

Intelligent verbal and action annotations have played an important role in railway stations, bus stands, police stations and road traffic to avoid crime diffusion, attacks and identify unhealthy human behavior. To predict and classify the human abnormal activities, which is occur at a remote location and/or unusual time by identify the movements from surveillance cameras for identifying the physical and verbal abnormal activities such as abuse, assault, arson, arrest, and fighting and raise an automatic alarm or notification. From the existing gap, an automation tool is required to detect the violence behavior with potential hazards and to prevent from crime to promote public safety. The proposed human activity and annotation recognition (HAAR) algorithm addresses the existing limitations by using contextual semantic analysis for identification of crowd, tracking of crowd behavior in large gatherings and people's abnormal behavior. It uses CNN and LSTM model to categorize the abnormal activity and generate a pattern by using contextual knowledge graph to identify the severity. HAAR model uses contextual knowledge graph with contextual vertices and edges to store human's verbal and physical action and objects involved. The captured attributes and its dependencies between the activities in a knowledge graph recognize the level of unusual behavior and generate ALERT messages to the authorities.

1. INTRODUCTION

Social abnormalities of human activity recognition are identified with an image or by surveillance video. Abnormal activity recognition is identified either by unusual objects or unusual scenes in the surveillance video or images. Action recognition is a high-level problem in surveillance video to identify the abnormalities in action and verbal [1] and also a challenge to detect, classify and predict abnormality in closed or crowded locations to prevent or minimize the occurrence of violence.

Normally, video-based human activity recognition system recognizes human activities in three phases. The sequence of frames from the input video is recognized in phase 1. In the second phase, extract the relevant features, and in the third, build a classifier to recognize the human activities [2]. If the videos consist of dense actions with many irrelevant frames, then it is very tough to analyze the content of a video stream to detect, track and recognize the objects, scenes, and violent activities by action or verbal. Violence may mitigate the psychological damages and particularly it is vulnerable to infants and victims of violence [3]. Hence, efficient and effective techniques are needed to identify potential hazards, prevent crimes, and promote public safety by detecting violent behavior with verbal action in the surveillance video content.

Furthermore, Human actions are identified by visual learners, knowledge mining, and action recognizers on individual frames by semantic contextual information [4].

Contextual information is viewed from two perspectives such as activities and human behavior processes. Location, identity, activity, time with user & role, process & task are used to find the contextual patterns. Frequent patterns are then used to classify the high-level activities by knowledge graph [5]. A knowledge graph supports the action and neighbor information to remove ambiguity and inconsistency in identifying the abnormalities [6]. Semantic knowledge graph is one of the efficient techniques to analyze abnormal human action (pose and pose let), attributes such as objects, verbal action and scene context [7].

Previous methods are mainly focused on the development of spatial-temporal representations to capture human activities at different speeds and various viewpoints on body parts interact [8]. Most approaches not recognize the complex actions due to their lack of explicit understanding between action and verbal activity [9].

To solve this issue, the proposed knowledge graph represents each action category with its attributes by semantic relation [10]. Deep learning models are used for analyzing abnormal activities and action recognition for inference into annotations.

2. LITERATURE SURVEY

Human Activity Recognition (HAR) is a rigorous task in analyzing and detecting human abnormal activity. They are

data collection, processing and activity classification [11]. Human abnormal activity recognition classifies the input data into activity categories. They are categorized into: (i) gestures of hand, leg, or body; (ii) actions; (iii) interaction of human-to-object or vice versa. (iv) Actions between the group; (v) individual behaviors; and (vi) event-based behavior [12] and it is shown in Figure 1.

Figure 1. Decomposition of human activities [12]

Human abnormal activity recognition is processed with activity, object detection, and knowledge-driven recognition. The complexity of situations often escalates due to various interactions between humans or between humans and environmental elements [13]. ANN, CNN and LSTM are used for human activity recognition. CNNs are used for image generation, object classification, image classification, hand-gesture recognition [14] but it needs pre-trained feature extractor in realistic conditions [15]. LSTM model extend with a squeeze and excitation (SE) model to recalibrate the different features to distinguish between similar activities by most informative data [16]. In the above models, data not semantically related.
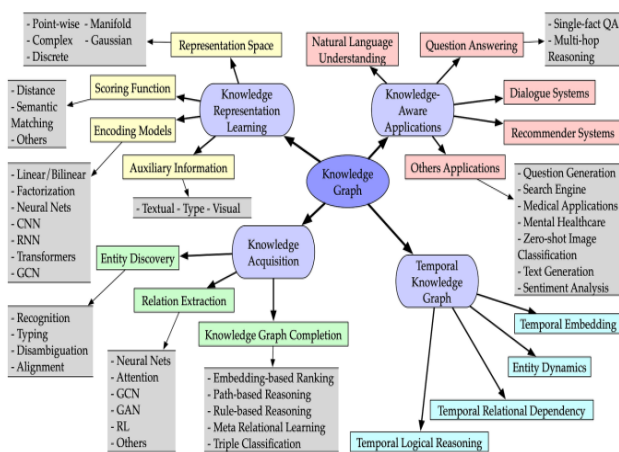

Figure 2. Categorization of research on knowledge graphs [17]

A framework to analyze and predict complex actions by a knowledge graph is required to identify verbal action and physical activity, objects in the situation for diagnosing behavioral challenges. Knowledge graphs consist of four categories. They are Knowledge Representation on Learning, Knowledge Acquisition, Temporal Knowledge Graphs and Knowledge-Aware Applications. Each category has its own enhanced functionality and it is shown in Figure 2.

In the temporal knowledge graph, predictive embedding model namely Contrastive Predictive Embedding (CPE) predict target entity embeddings. CPE uses knowledge graph for predicting the target embedding directly and leveraging the similarity between the target and predicted embeddings [18]. But video captioning is extremely tough in knowledge graphs [19]. In a real time, surveillance monitoring system, a knowledge graph with an adaptive prioritization mechanism is helpful for tracking the human's abnormal behaviors [20, 21]. Prioritization requires tokenization for detecting salient objects, video frames and speech transcripts [22]. The predefined tasks listed earlier to ensure contextual and semantic interpretation from videos render appropriate captions to enhance understandability.

Caption or annotation is required for object detection and instance segmentation in video frames [23]. For action recognition, fuse objects and human interaction by graph-based recognition support contextual information such as image scenes, objects, and their interactions without semantic support [24].

Hence, a model is required to relate an individual human activity, group activities and their actions by semantically related graph, called contextual knowledge graph which is helpful in predicting and identifying abnormal activities. A proposed human activity and annotation recognition (HAAR) use contextual knowledge graph with entities, attributes and it relationships for semantic descriptions with alarm based on the subject-predicate for verbal and physical abnormality.

3. METHODOLOGY

The proposed human activity and annotation recognition (HAAR) uses knowledge graph to identify the humans' verbal and physical activities. Surveillance video of moving and still images such as standing, walking, running, jumping, and going up/down stairs are identified for contextual learning. It uses predefined templates for assault, arson, arrest, and fighting and abuse activity to identify the context of action in the surveillance video and stored as a node in the knowledge graph.

Human activity and annotation recognition (HAAR) uses Contextual Knowledge Graph (CKG) which consists of knowledge vertex (Kv) and knowledge edges (Ke). Knowledge vertex (Kv) stores human activity, verbal action and abnormal objects and vertex are connected by the knowledge edges (Ke) such as area, zone and time and it is represented as CKG = (Kv, Ke).

Human activity and verbal recognition help to infer and predict human verbal annotations. The following algorithm explains the proposed human activity and annotation recognition (HAAR) in steps:
Algorithm for identifying human activity and annotation recognition (HAAR):
1. *Collect abnormal activity images*
2. *Classify abnormal activities types*
3. *Analyze the abnormal activities style*
4. *Analyze intensity level*
5. *Analyze abnormal activity function as funny abnormal activity, bitter or embarrassed abnormal activity, self-conscious abnormal activities, amused abnormal activities, mirthful*

abnormal activates, untrue abnormal activities by facial expression, head and upper body movement

6. *Develop a model to connect dialogue with gesture motion*
7. *Discover the gesture happening combined with dialogue act*
8. *Classify and rank the tokens into category*
9. *Generate Alert for the verbal and physical annotations based on the Subject-Predicate-ALERT*

Step 1: Collect abnormal activity images/Data

The way that section titles and other analysis the abnormalities by contextual data, places of surveillance video data, type of Data, Data acquisitions and data format is required. Data is collected from various sources such as surveillance cameras and microphones from crowded or remote public places for identifying contextual action recognition and it is shown in Table 1.

Table 1. Data for abnormal activity recognition

Contextual Data	Face Recognition, Action Recognition, Verbal Recognition
Type of Data	Videos sequences (Frames) and face images for verbal notification
Data Acquisitions	Publicly available videos captured from multiple cameras installed in public places, crowded places, large gatherings, remote places
Data Format	PNG (Portable Network Graphic) files
Experimental Factors	Indoor/outdoor scenes of large gatherings, various object types, crowd density
Experimental Features	Number of frames, number of clips, Variable video clip lengths
Abnormal Activity Template	The dataset contains action images

Step 2: Classify abnormal activities types

The identification of activities that deviate from anticipated behavior is known as abnormal activity. It is identified from body activity and verbal action on input data. Types of violence is explained in Figure 3 and in Table 2.
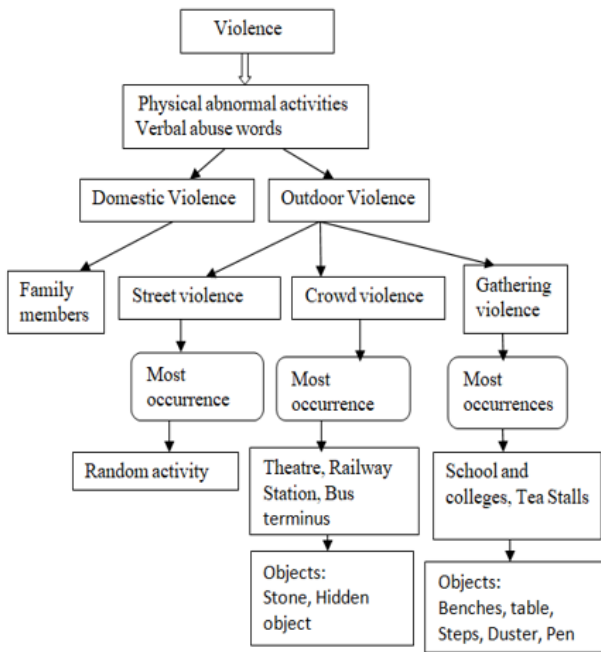

Figure 3. Abnormal activities types

Table 2. Abnormal activities and their attributes

No.	Entity	Attributes
1	Assault	Scratching, pushing, physically restraining, throwing things, etc.
2	Arson	Disorder, retardation, etc.
3	Arrest	Observe, move, surrender, capture
4	Fighting	Speed, agility, power, balance, coordination, reaction time, etc.
5	Abuse	Hitting, shaking, burning, pinching, biting, choking, etc.

Step 3: Analyze the abnormal activities style by contextual knowledge graph

Human activity and annotation recognition (HAAR) use Contextual Knowledge Graph (CKG) represented as CKG = (Kv, Ke), Kv as knowledge vertex and Ke for knowledge edges. Knowledge vertex (Kv) stores human activity, verbal action and abnormal objects. Vertices are connected by the knowledge edges (Ke) such as area, zone and time. A contextual knowledge graph captures the relationships and dependencies between different activities for activity recognition, behavior analysis, and personalized recommendations. Figure 4 shows the semantic view of the contextual knowledge graph.
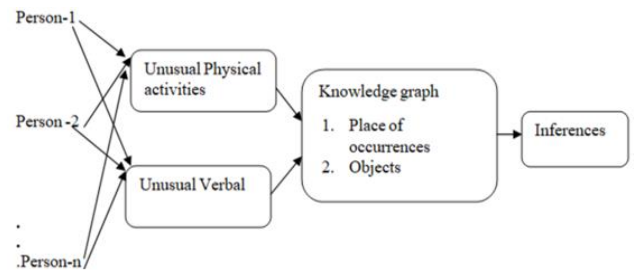

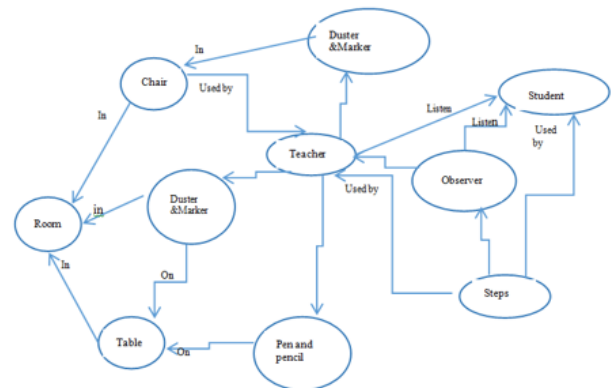**Figure 4.** Overview of contextual knowledge graph

Figure 5. Semantic view of knowledge graph in a class room

A contextual knowledge graph for human activities and their complex relationships and dependencies in a class room scenario is shown in Figure 5 and it is explained with knowledge vertices, edges and its relationship such as spatial, observational and actions.

Knowledge graph nodes (Kv):

Indoor physical objects: Room, Door, Chair, Table, Floor

People: student, Teacher, Parent, Observer

Activities: Normal, abnormal

Knowledge graph relativity (Ke):

There are 3 relationships namely spatial, observational, and actions.

(i) Spatial relationships (in-door):

R_Near: Proximity (e.g., Room-Door, Steps)

R_On: Physical placement (e.g., Chair, table, Duster, pen, pencil, water bottle)

R_In: Containment (e.g., Room-Table)

(ii) Observational relationships:

R_Watch: Watching/observing (e.g., Student-Activity, teacher-Activity)

(iii). Action relationships:

R_Execute: Performing actions (e.g., Teacher, Observer)

The resulting knowledge graph can be used to train machine learning models to recognize and classify human activities.

Step 4 & 5: Analyze abnormal activity function as funny abnormal activity, bitter or embarrassed abnormal activity, self-conscious abnormal activities, amused abnormal activities, mirthful abnormal activates, untrue abnormal activities

To formulate and train a knowledge graph for human abnormal activities such as abuse, assault, arson, arrest, and fighting, by clustering and embedding visualization techniques to organize different types of activities and the factors that may contribute to them. They are

1. Contextualizing abnormal behavior: An edge on a knowledge graph represents contextual information such as time, location, and social relationships to identify the abnormal behavior.

2. Identifying correlations: A knowledge graph can help in identifying correlations between different events and behaviors to identify the severity of abnormal behavior.

Step 6: Develop a model to connect verbal with gesture motion

To train a knowledge graph by both verbal and action at the time of abuse, assault, arson, arrest, and fighting involves following steps.

Identify relevant entities and relationships: Identify the entities (people, objects, and events) that are relevant to the abnormal activities and the relationships between them. For example, for the activity of "arson", relevant entities might include buildings, accelerants, ignition sources, and people who were in the vicinity at the time of the fire. The relationships between these entities might include "cause", "location", and "time".

The following mathematical formula shows the representations of verbal and physical abnormalities:

Let, Eq. (1) explain on contextual knowledge graph with vertex and edges (CKG(Kv,Ke)), V- Verbal abnormalities, W- Physical abnormalities, $F(AE)$- List of abnormalities, Q- A abnormality event.

$$CKG(Kv,Ke) = \sum (V,W) \; \varepsilon \; F(AE) \| Q(V,W) \tag{1}$$

$$\sum_{v,w \in F(AE)} Q(v,w) \| E(v,:) - E(w,:) \tag{2}$$

$F(AE)$- List of abnormalities, V- Verbal abnormalities, W- Physical abnormalities, Q- A abnormality event.

$$\sum_{v,w=1}^{F^{E(x)}} \alpha . Q(v,w) . (E(v,x) - E(w,x)) \tag{3}$$

X is entity of an event (E).

$$\sum_{j=1}^{a} \sum_{v,w=1}^{F^{E(x)}} Q(v,w) . [E^2(i,j) - E(v,x).E(w,x)] \tag{4}$$

Object involved during abnormalities, j- Location of abnormalities.

$$OSP(e_1, e_h) = \frac{\sum_{x=1}^{b} a_{ix} . a_{jx}}{\sqrt{\sum_{x=1}^{b} a_{wx}^2} . \sqrt{\sum_{x=1}^{b} a_{wx} . 2_x}} \tag{5}$$

OSP- Opinion similarity on physical activity.

b= Leg and hand physical activity, e_l- Entity on leg, e_h- Entity on hand.

$$OSV(e_1, e_h) = \frac{\sum_{x=1}^{b} a_{ix} . a_{jx}}{\sqrt{\sum_{x=1}^{b} a_{wx}^2} . \sqrt{\sum_{x=1}^{b} a_{wx} . 2_x}} \tag{6}$$

OSV- Opinion similarity on verbal activity, f = Facial expression, v= Verbal activity.

Step 7: Discover the gesture happening combined with verbal act

Analysis and annotate verbal and physical activity data: To build a knowledge graph annotating physical and verbal data by the following Figure 6.
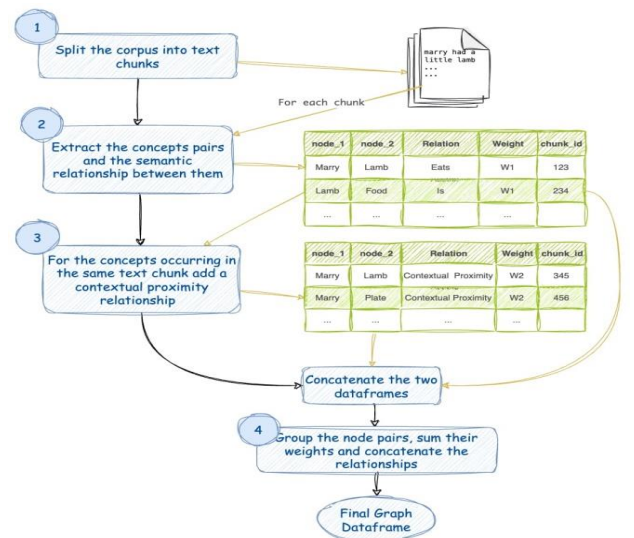

Figure 6. Grouping of abnormalities

Step 8: Generate, classify and rank tokens for the verbal and physical annotations

Critical level of the violence is analyzed by its severity. Level 1 is assigned to physical abnormalities. Level 2 is for verbal and level 3 is for when both physical and verbal activity is identified.

$$R(v,w) = OSP(e_1, e_h) \ if \ v \neq w = R1$$
$$R(v,w) = OSV(r_f, r_v) \ if \ v \neq w = R2 \quad (7)$$
$$R(v,w) = \{OSP(e_1, e_h) U \{OSV(r_f, r_v)\} = R3$$

Step 9: Generate alert for the verbal and physical activity by Subject-Predicate-ALERT annotations

Annotate the data: Annotating the data involves adding metadata to the data that helps to describe the entities and their relationships in the form of subject-predicate and alert. This can include adding labels, descriptions, and identifiers to the data. It is shown in Table 3 in the surveillance video in the same zone and time. Based on verbal and physical activities, annotations are generated for alert messages.

Table 3. Subject-Predicate-ALERT annotations

No.	Subject (Node)	Predicate (Edges)	Alert
1	Person 1	Abuse, physical action	Nearest police station
2	Person 2	Physical action Knife	

4. RESULT

Real time surveillance data or images are analyzed to identify the abnormal physical activities of human and the objects present in the scenario. Figures 7 and 8 show human's abnormal physical activity in the common place.

Figure 7. Sample abnormal physical activity

Figure 8. Sample abnormal activity

Some predefined verbal words are used in the verbal abnormality. Figures 9-13 show the verbal words used during abnormal human activities.
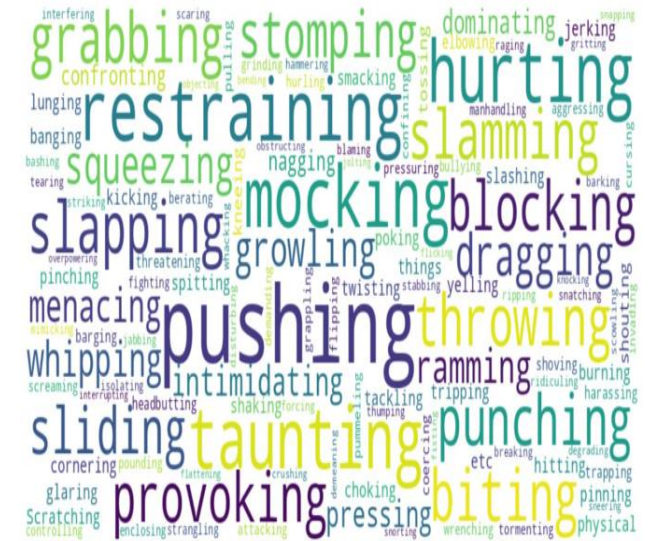

Figure 9. Word cloud of arson

Figure 10. Word cloud of assault

Figure 11. Word cloud of arrest

Figure 12. Word cloud of abuse

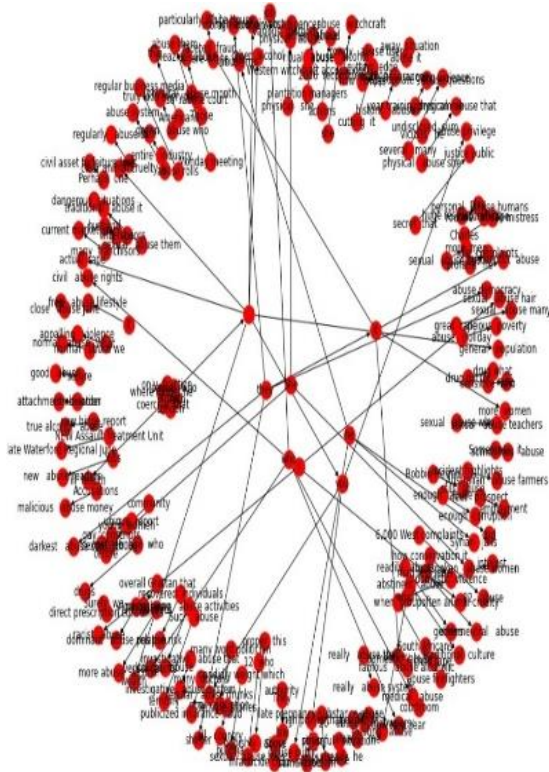**Figure 13.** Word cloud of fighting

Figure 14. Knowledge graph construction of abnormal activity

A contextual knowledge graph uses nodes and edges to support the abnormal activity recognition and classify its level from verbal and physical activity of human behavior, as shown in Figure 14.

Table 4 and Table 5 show the performance matrices such as precision, recall, F1-Score and accuracy of abnormal behaviors such as abuse, arrest, arson, assault, and fighting are calculated with and without the knowledge graph.

Figure 15 and Figure 16 show the precision, recall, accuracy of proposed HAAR model without and with knowledge graph. Figure 17 shows the confusion of the matrix of HAAR model. It shows the accuracy on abnormal activity data.
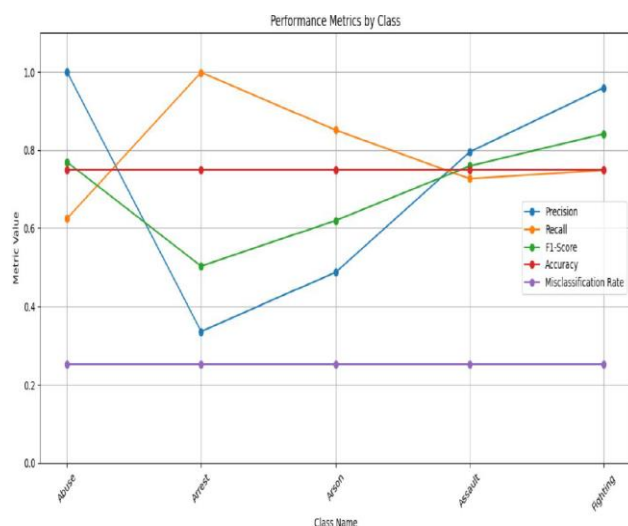

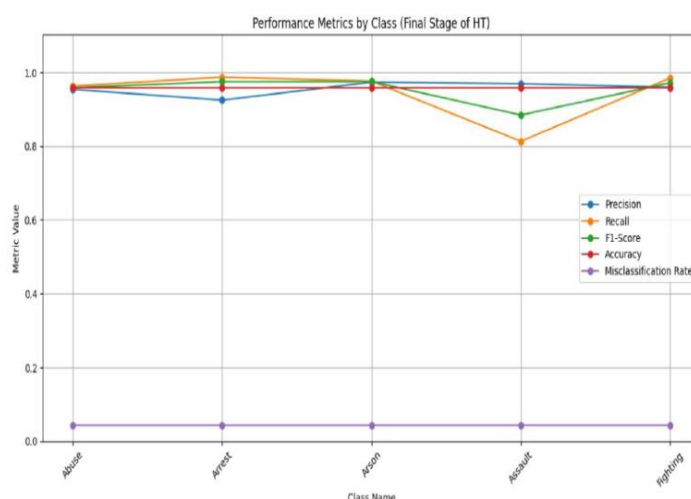**Figure 15.** Matrices of HAAR model without knowledge graph

Figure 16. Matrices of HAAR model with knowledge graph

Table 4. Performance matrices of the proposed HAAR model on without using the knowledge

Class Name	Precision	1-Precision	Recall	1-Recall	F1-Score	Accuracy	Misclassification Rate
Abuse	1.0000	0.0000	0.6247	0.3753	0.7690		
Arrest	0.3358	0.6642	0.9989	0.0011	0.5026	0.7486	0.2514
Arson	0.4868	0.5132	0.8510	0.1490	0.6194		
Assault	0.7950	0.2050	0.7265	0.2735	0.7592		
Fighting	0.9597	0.0403	0.7487	0.2513	0.8412		

Table 5. Performance matrices of the proposed HAAR model at the end of abnormal event

Class Name	Precision	1-Precision	Recall	1-Recall	F1-Score	Accuracy	Misclassification Rate
Abuse	0.943	0.043	0.960	0.0370	0.9586		
Arrest	0.926	0.0750	0.9873	0.0127	0.9751	0.9581	0.0419
Arson	0.9737	0.0263	0.9765	0.0225	0.9751		
Assault	0.9696	0.0304	0.8133	0.1867	0.8846		
Fighting	0.9597	0.0403	0.9850	0.0150	0.9722		

Figure 17. Confusion matrix of HAAR

5. CONCLUSION

The contextual knowledge graph stores verbal activity and physical motions on its vertices and edges. The contextual knowledge graph represents events, resources/objects, and locations on its knowledge vertex, where the edges relate activities/cooperation and relationships between human, location and objects. The activities performed by the human include the evaluation of physical movements, verbal usages, and other conditions such as objects and locations for activity recognition. Surveillance video data is converted into a knowledge graph for automation in abnormalities and sends alert messages by exploring data and processes. The proposed model generalized the abnormal activities such as abuse, arrest, arson, assault and fighting class labels with 95%, 92%, 97%, 97%, and 96% accuracy respectively. Finally, the model has achieved a high testing accuracy of 96% with just a 4% misclassification rate.

REFERENCES

[1] Huang, H.B., Zheng, Y.L., Hu, Z.Y. (2024). Video abnormal action recognition based on multimodal heterogeneous transfer learning. Advances in Multimedia, 2024: 4187991. https://doi.org/10.1155/2024/4187991

[2] Dwivedi, N., Singh, D.K., Kushwaha, D.S. (2023). A novel approach for suspicious activity detection with deep learning. Multimedia Tools and Applications, 82: 32397-32420. https://doi.org/10.1007/s11042-023-14445-7

[3] Kumar, M., Patel, A.K., Biswas, M., Sengar, S.S. (2024). Human abnormal activity recognition from video using motion tracking. International Journal of Image, Graphics and Signal Processing, 16: 52-63. https://doi.org/10.5815/ijigsp.2024.03.05

[4] Elharrouss, O., Almaadeed, N., Al-Maadeed, S. (2021). A review of video surveillance systems. Journal of Visual Communication and Image Representation, 77: 103116. https://doi.org/10.1016/j.jvcir.2021.103116

[5] Khowaja, S.A., Yahya, B.N., Lee, S.L. (2020). CAPHAR: context-aware personalized human activity recognition using associative learning in smart environments. Human Centric Computing and Information Sciences, 10: 35. https://doi.org/10.1186/s13673-020-00240-y

[6] Sadasivan, A., Gananathan, K., Pal Nesamony Rose Mary, J.D., Balasubramanian, S. (2025). A systematic survey of graph convolutional networks for artificial intelligence applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 15(2): e70012. https://doi.org/10.1002/widm.70012

[7] Beddiar, D.R., Nini, B., Sabokrou, M. (2020). Vision-based human activity recognition: A survey. Multimedia Tools and Applications, 79: 30509-30555. https://doi.org/10.1007/s11042-020-09004-3

[8] Vrigkas, M., Nikou, C., Kakadiaris, I.A. (2015). A review of human activity recognition methods. Frontiers in Robotics and AI, 2: 160288. https://doi.org/10.3389/frobt.2015.00028

[9] Lindwall, O., Mondada, L. (2025). Sequence organization in the instruction of embodied activities. Language and Communication, 100: 11-24. https://doi.org/10.1016/j.langcom.2024.11.003

[10] Wu, P.C., Tu, H.B., Mou, X., Gong, L.H. (2025). An intelligent energy management method for the manufacturing systems using the knowledge graph and large language model. Journal of Intelligent Manufacturing, 1-20. https://doi.org/10.1007/s10845-025-02587-4

[11] Jlidi, N., Jemai, O., Bouchrika, T. (2025). Enhancing human action recognition through transfer learning and body articulation analysis. Circuits Systems and Signal Processing, 1-29. https://doi.org/10.1007/s00034-025-03026-8

[12] Alsaadi, M., Keshta, I., Ramesh, J.V., Nimma, D., Shabaz, M., Pathak, N., Singh, P.P., Kiyosov, S., Soni, M. (2025). Logical reasoning for human activity recognition based on multisource data from wearable device. Scientific Reports, 15: 380. https://doi.org/10.1038/s41598-024-84532-8

[13] Kumar, M., Patel, A.K., Biswas, M. (2024). Real-time detection of abnormal human activity using deep learning and temporal attention mechanism in video surveillance. Multimedia Tools and Applications, 83: 55981-55997. https://doi.org/10.1007/s11042-023-17748-x

[14] Raj, R., Kos, A. (2023). An improved human activity recognition technique based on convolutional neural network. Scientific Reports, 13: 22581. https://doi.org/10.1038/s41598-023-49739-1

[15] Cruciani, F., Vafeiadis, A., Nugent, C. (2020). Feature learning for Human Activity Recognition using Convolutional Neural Networks. CCF Transactions on Pervasive Computing and Interaction, 2: 18-32. https://doi.org/10.1007/s42486-020-00026-2

[16] Khan, M., Hossni, Y. (2025). A comparative analysis of LSTM models aided with attention and squeeze and excitation blocks for activity recognition. Scientific Reports, 15: 3858. https://doi.org/10.1038/s41598-025-88378-6

[17] Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S. (2020). A survey on knowledge graphs: Representation, acquisition and applications. IEEE Transactions on Neural Networks and Learning Systems, 33(2): 494-514. https://doi.org/10.1109/TNNLS.2021.3070843

[18] Liu, C., Wei, Z.H., Zhou, L.X. (2025). Contrastive Predictive Embedding for learning and inference in knowledge graph. Knowledge-Based Systems, 307: 112730. https://doi.org/10.1016/j.knosys.2024.112730

[19] Wajid, M.S., TerashimaMarin, H., Najafirad, P., Wajid, M.A. (2024). Deep learning and knowledge graph for image/video captioning: A review of datasets, evaluation metrics, and methods. Engineering Reports, 6: e12785. https://doi.org/10.1002/eng2.12785

[20] Pham, V.H., Nguyen, Q.H., Le, T.T., Nguyen, T.X.D., Phan, T.T.K. (2022). A proposal model using deep learning model integrated with knowledge graph for monitoring human behavior in forest protection. TELKOMNIKA, 20(6): 1276-1287. http://doi.org/10.12928/telkomnika.v20i6.24087

[21] Singhal, A. (2012). Introducing the knowledge graph: Things, not strings. Google Official Blog. https://blog.google/products/search/introducing-knowledgegraph-things-not, accessed on May 16, 2012.

[22] Gu, X., Chen, G., Wang, Y.F., Zhang, L.B., Luo, T.J., Wen, L.Y. (2023). Text with knowledge graph augmented transformer for video captioning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, pp. 18941-18951.

[23] Zhong, M.S., Zhang, H., Xiong, H., Chen, Y.D., Wang, M.W., Zhou, X.Y. (2022). Kgvideo: A video captioning method based on object detection and knowledge graph. Computer Vision and Image Understanding, 1-7. https://doi.org/10.2139/ssrn.4017055

[24] Sabbu, S., Ganesan, V. (2021). LSTM-based neural network to recognize human activities using deep learning techniques. Applied Computational Intelligence and Soft Computing, 2022: 1681096. https://doi.org/10.1155/2022/1681096