Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 7, 2021, Pages. 1330 - 1340
Received 05 May 2021; Accepted 01 June 2021.

# Integrated Algorithm for Human Activity Recognition using GAN

**B.Mounika[1], T.Srinivasa Rao[2], K. Vijaya Naga Valli[3],**

[1,2,3]Assistant Professor, Department of CSE, SRKR Engineering College (A),

bmounika@srkrec.ac.in, tsrao@srkrec.ac.in, kvnv@srkrec.ac.in

**Abstract:** Activity recognition is one of the significant task and most difficult tasks in computer science. Several actions are performed by the various people and the activities are collected based on the actions performed by the people. Many existing approaches are developed to recognize the several individual actions, pair-wise interactions, and pose which is not an easy task. Deep Learning (DL) is most widely used to solve the various issues in human activity recognition. In this paper, the Enhanced algorithm for human activity recognition is developed with the integration of Generative Adversarial Network (GAN) with HaaR features with bounding box. This is a very efficient method that will provide the individual activities and also the group activities, actions that are recognized by the proposed approach. The performance is calculated by using parameters such as accuracy. The comparative results are CERN, CAR, and Enhanced GAN.

**Keywords**: GAN, HaaR, CERN, CAR.

## Introduction

From the past many years, recognition of human activities become more complex to the normal cameras and existing algorithms. Understanding the complex videos and analyzing the content having huge demand. Human activities can be analyzed by the many existing algorithms to recognize the accurate activity in many videos and also in images. Every year, human activity recognition becomes a more trending topic for research.

Human activity becomes more complicated in various levels. This work is mainly divided into three different levels based on the complexity: single action, group activities and behavior of the person in crowd. In figure 1, various levels are shown. In the first level, the single action is considered as the single person action which is having the human pose and the motions of the human bodies are present with biased data. The second level is based on the activities of human present in the crowd. It is very difficult to obtained the actual information from the individual person. The main aim of this research is to provide the accurate information about the abnormal activities that are identified in the motion videos.

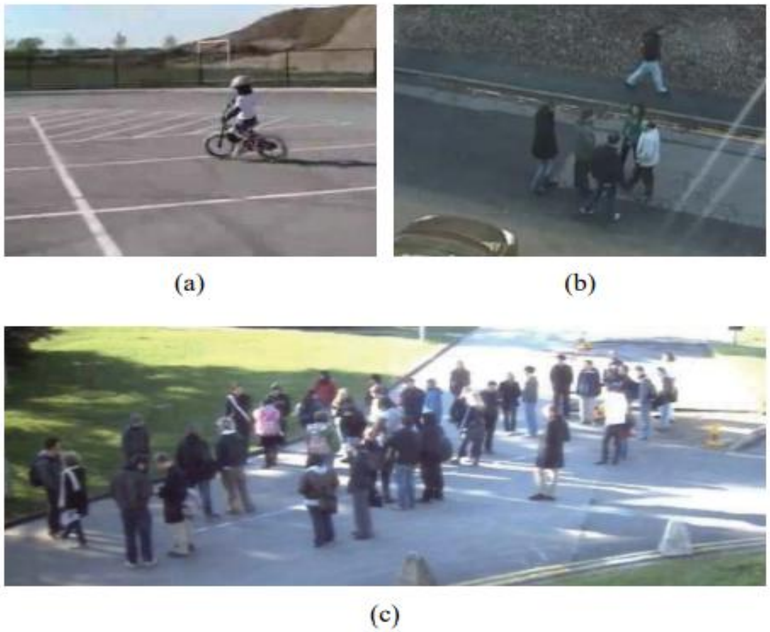

**Figure: 1 three levels of human activity analysis: (a) Human action; (b) Group activity; (c) Crowd behavior.**

This paper is mainly focused on recognizing the human activity by using the integration of GAN with HaaR features. This will increase the accuracy of recognition by using the HaaR features. The proposed approach is also used to recognize the multi-person detection and also the multi-person tracking and activity recognition.

## Literature Survey

In this section several activity recognition based algorithms are discussed.

A.Behra.et. al introduced a methodology for perceiving exercises utilizing video from an egocentric (first-individual view) arrangement. Their methodology surmises action from the cooperations of items and hands [1]. Rather than past ways to deal with action acknowledgment, they didn't need to utilize a transitional, for example, object identification, present assessment, and so on Demonstrating the spatial dissemination of visual words compared to nearby highlights additionally worked on the presentation of action acknowledgment utilizing the pack of-visual words portrayal. A.Fathi.et.al introduced a technique to break down day-by-day exercises [2], like feast readiness, utilizing video from an egocentric camera. Their technique performed deduction about exercises, activities, hands, and articles.

Day-by-day activities are a difficult space for recognition of activities that were appropriate to an egocentric methodology. They presented an original portrayal of activities dependent on object-hand associations and tentatively exhibited the unrivaled presentation of our portrayal in contrast with standard movement portrayals like a sack of words. H. Bay. et.al introduced a clever scale-and pivot invariant interest point locator and descriptor, begat SURF (Speedup Robust Features). It approximated or even outflanked recently proposed plans concerning repeatability, uniqueness, and strength, yet can be registered and looked at a lot quicker [3]. This was accomplished by depending on fundamental pictures for picture convolutions; by expanding on the qualities of the main existing indicators and descriptors (if, utilizing a Hessian framework based measure for the finder, and dispersion based descriptor); this lead to a mix of novel location, depiction, and coordinating with steps. A.Behera et al., introduced a technique for ongoing checking of work processes in an obliged climate. The observing framework ought not exclusively to have the option to perceive the current advance yet additionally give directions about the conceivable subsequent stages in a continuous work process [4]. They resolved this issue by utilizing a strong methodology (HMM-pLSA) which depended on a Hidden Markov Model (HMM) and generative model like probabilistic Latent Semantic Analysis (pLSA). The quantization and classifier were both arranged in an earlier taking-in stage from preparing information. A movement was addressed by a Markov model over nuclear occasions. P.Matikainen et al., depicted the well-known pack of words worldview for activity acknowledgment errands depended on building histograms of quantized elements, ordinarily at the expense of disposing of all data about connections between them [5]. They proposed a straightforward and computationally effective strategy for communicating pairwise connections between quantized elements that consolidated the force of discriminative portrayals with key parts of Naive Bayes.

E.Shechtman et al., [6] introduced a methodology for estimating similitude between visual elements (pictures or recordings) because of coordinating with interior self likenesses. The interior self-similitudes were proficiently caught by a conservative neighborhood "self closeness descriptor", estimated thickly all through the picture/video, at various scales, while representing nearby and worldwide mathematical twists. They contrasted our action with usually utilized picture-based and video-based closeness gauges and exhibited its materialness to protest identification, recovery, and activity detection. W.Liu et al., portrayed Increasing the expressiveness of subjective spatial calculi was a fundamental stage towards meeting the necessities of uses [7]. They consolidated probably the most popular calculi in subjective spatial thinking (QSR), the RCC8 polynomial math for addressing topological data, and the Rectangle Algebra (RA) and the Cardinal Direction Calculus (CDC) for directional data. Lothey et al., proposed another model portrayal that had a less prohibitive earlier on the math and number of nearby highlights, where the calculation of every neighborhood include was impacted by its k nearest neighbors and models might contain many elements and a clever solo online learning calculation that was fit for assessing the model boundaries productively and precisely [8].

Lothey et al., introduced a technique for extricating particular invariant highlights from pictures that could be utilized to perform solid coordinating between various perspectives on an article or scene [9]. This way to deal with acknowledgment could heartily distinguish articles among mess and impediment while accomplishing close to ongoing execution. A.Fathi.et.al resolved the issue of taking in object models from egocentric video of family exercises, utilizing incredibly feeble management [10]. By utilizing Multiple Instance Learning to coordinate with object examples across groupings, they found and confined item events. Article portrayals are refined through transduction and item level classifiers were prepared. They exhibited empowering brings about identifying novel item occurrences utilizing models delivered by feebly administered learning.

By and large, customary AI dependent on manual element extraction is the most normal strategy for perceiving human action. Profound learning models with programmed include extraction are additionally generally adjusted to this local area [11, 12, 13,].

In general, regular ML approaches depend on the accessibility of adequate amounts of test information [14,15,16]. In this way, little example issue seriously restricts the expected commitment of human action acknowledgment advances. Truth be told, most earlier works on human movement acknowledgment zeroed in on the characterization exhibitions and tried to ignore the significance of getting ready and creating ideal sensor information. To additionally work on the exhibition of HAR, productive information age techniques ought to be explored. In this way, the investigation is done based on GANs structure which is an interesting issue in DL local area[17].

## A Generative Adversarial Network (GAN)

Generative Adversarial Networks (GAN) is the DL technique like convolutional neural networks (CNN). GAN is an unsupervised learning task present in the ML that involves finding the regular patterns in the given input data which is the model used to generate the new samples that specious could have been drawn from the original dataset.

GAN is a very significant technique that trains a generative model by solving the supervised learning problem by using two sub-models: The proposed model gives the huge training that creates the new samples and the differentiator model which classifies the samples with real (from the domain) or fakes (generated). These models are more trained based on a zero-sum game which gives better training on the given dataset. GAN is the existing and fast-developing field that gives real-time samples from a wide range of issue domains, the image conversion process changes the conversion images of summer to winter that generates the image realistic images of objects, scenes that no one can identify as fake.

- This architecture is dynamic in nature which gives the improved training with this model and solves the unsupervised issue which is used in both generative and a discriminative model.
- GAN provides the specific enlightened data augmentation to solve the various issues that needs a generative output which is called as image to image conversion.
- The main aim of this paper is to provide the accurate detection and recognition of human activities in the given videos as input.

Figure 2 explains the process of the proposed algorithm that involves a few steps. The training is given with pre-trained model VGG-16 is used to extract the accurate features of human activity recognition. VGG-16 is the deep learning model that gives the best training to several domains. VGG-16 is the one of the CNN architecture that improves the analysis to get the increased output. This network uses the small 3 x 3 filters and consists of several layers such as connected layers and pooling layers.

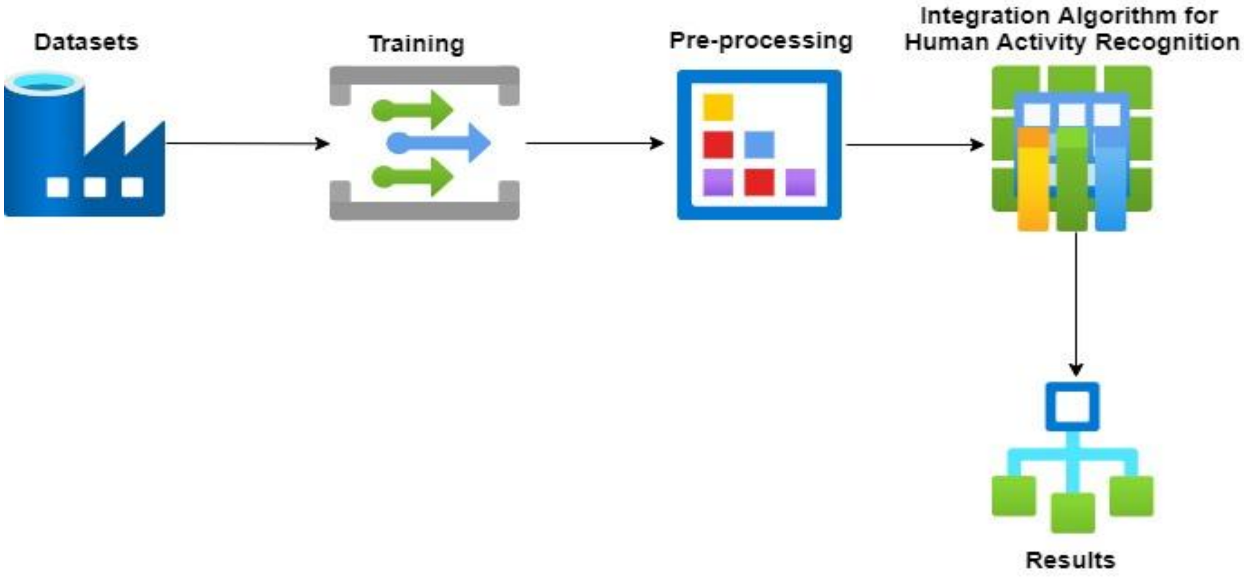

**Figure 2: System Architecture**

## Generative Adversarial Networks in HAAR

The traditional GAN system is developed and generator G(z) and the discriminator D(x) and 'z' represent the random noise. The generator G(z) attempts to create an ever-increasing number of sensitive information to 'fools' the discriminator D(x), while the discriminator D(x) expects to distinguish the phony information from the genuine information. These two antagonistic rivals are advanced to overwhelm one another and play a lose-lose situation (likewise called the min-max game) in the entire preparing process. The arbitrary commotions z ∈ R N (generally typical dispersion or gaussian appropriation) are given as the contribution of the generator G(z). And afterward, the generator G(z) will create manufactured information, x˜ = G(z). The genuine information x and phony information x˜ will be both taken care of to the discriminator D(x), and afterward, the discriminator D(x) will yield a scalar which addresses the likelihood of information are from the genuine information dispersion P(x) rather than the generator G(z). The two ill-disposed players are advanced by the antagonistic preparing process. The worth capacity of this ill-disposed cycle is as per the following (GANs gain proficiency with the generator G(z) and the discriminator D(x) by taking care of Nash balance issue):

$$\min_{G} \max_{D} V\ (D, G) = E_{x\sim Pdata\ (x)}[\log D(x)] +\ E_{z\sim P_z(z)}[\log(1 - D\big(G(z)\big))]$$

Received 05 May 2021; Accepted 01 June 2021.

Where Pz(z) represents the random noises (uniform distribution in most GANs at the early phase, Pz(z) = U(0, 1). The generator and discriminator is represented as G(z) , D(x) in original GANs[7] which is developed multilayer perceptrons. The training is done by using stochastic gradient descent (SGD) represents the Equation 1.

## Simulation Results

The implementation is done by using MATLAB. To show the effective simulation results the comparison between various algorithms is used to show the performance. MATLAB has powerful libraries that can implement the proposed algorithm very efficiently. The datasets that are used for experiments is volleyball dataset and pedestrian dataset. The dataset volleyball dataset consists of 1000 images and pedestrian dataset consists of 1000 images. For training 800 images and for testing 200 images.

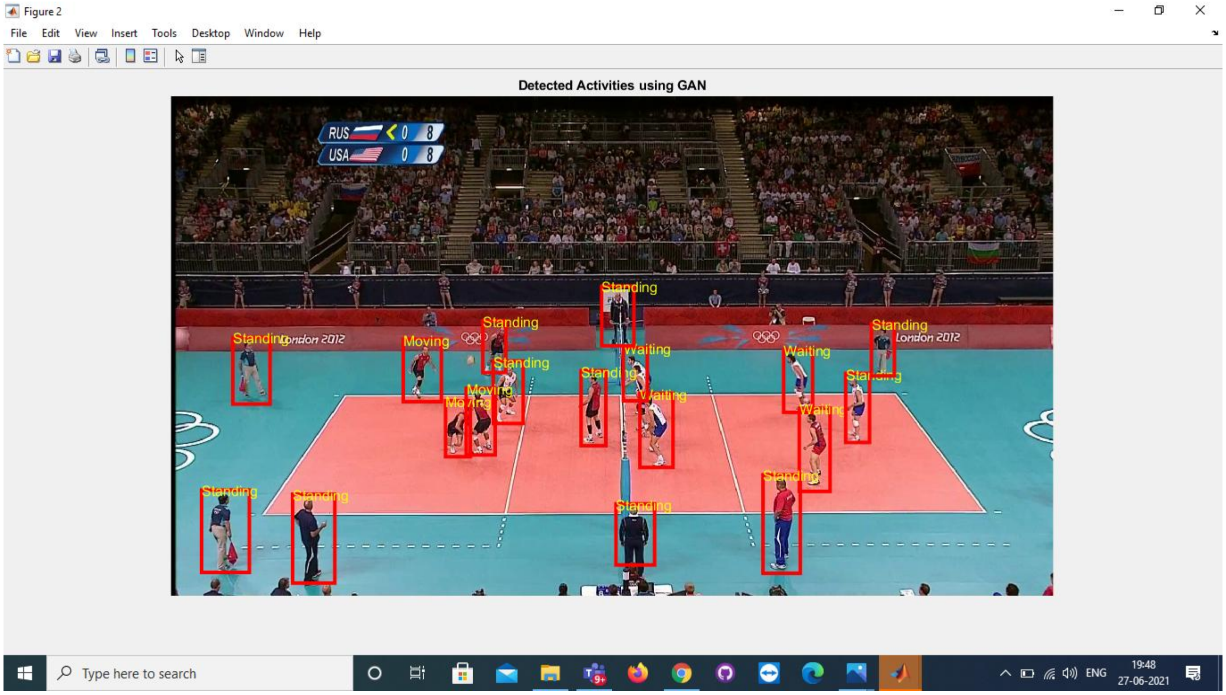

**Figure 2: Showing the human activities by using GAN on Volleyball dataset**

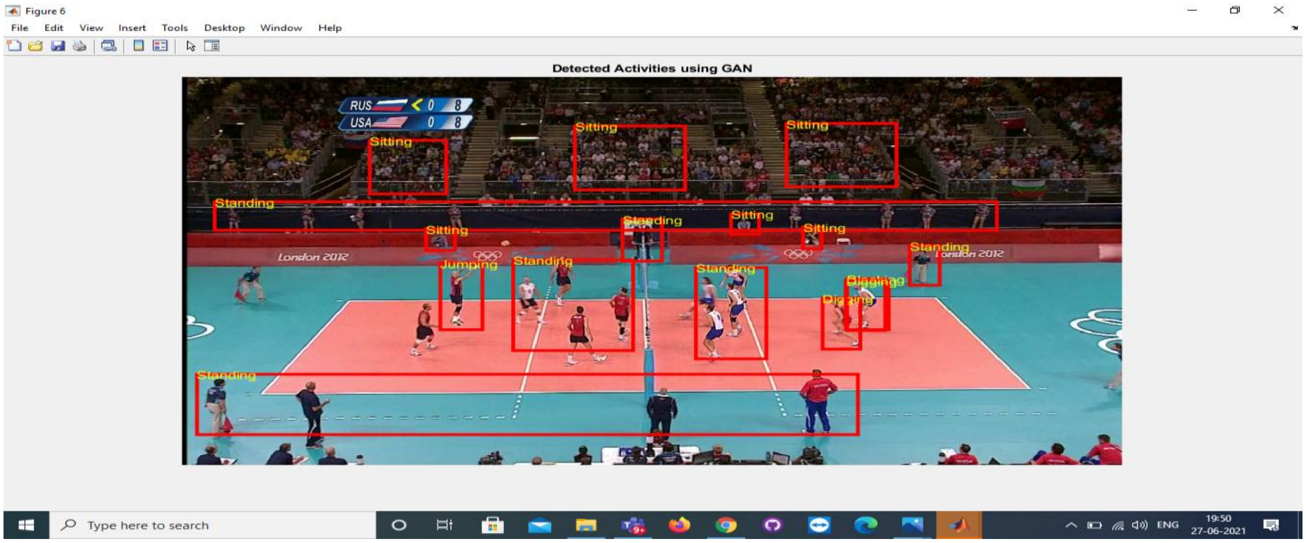

**Figure 3: Showing the human activities by using GAN on Volleyball dataset by integrating the bounding box.**

**Table 1: Shows the performance of Existing and Proposed Algorithms**

| Algorithm | Sensitivity | Specificity | Accuracy | F1-Score | Duration (MS) |
|---|---|---|---|---|---|
| CERN | 53.43 | 52.12 | 49.10 | 46.78 | 12.43 |
| CAR | 75.32 | 76.45 | 84.56 | 56.98 | 7.87 |
| Integrated GAN | 88.54 | 88.32 | 87.98 | 61.32 | 4.12 |

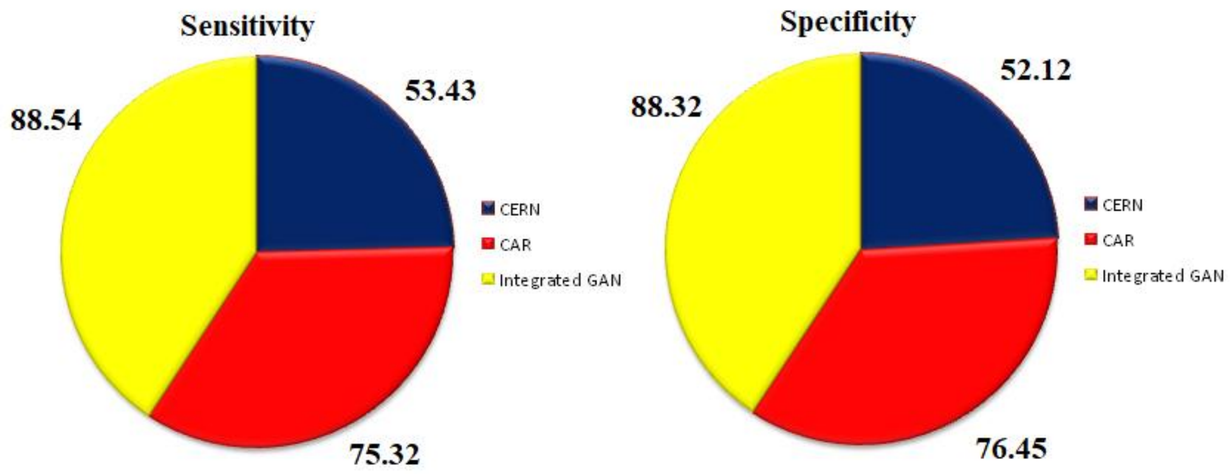

**Figure 4: Shows the Performance based on Sensitivity and Specificity**

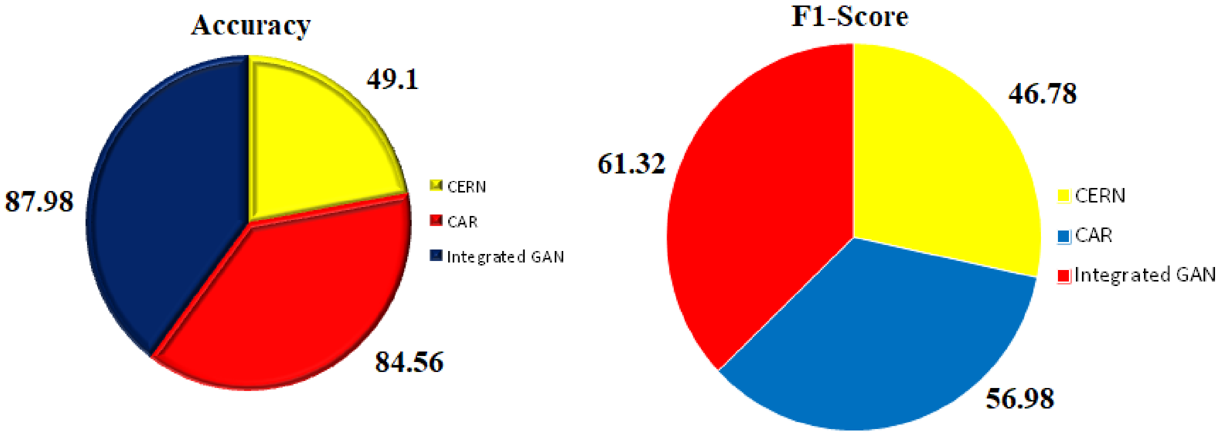

**Figure 5: Shows the Accuracy and F1-Score**

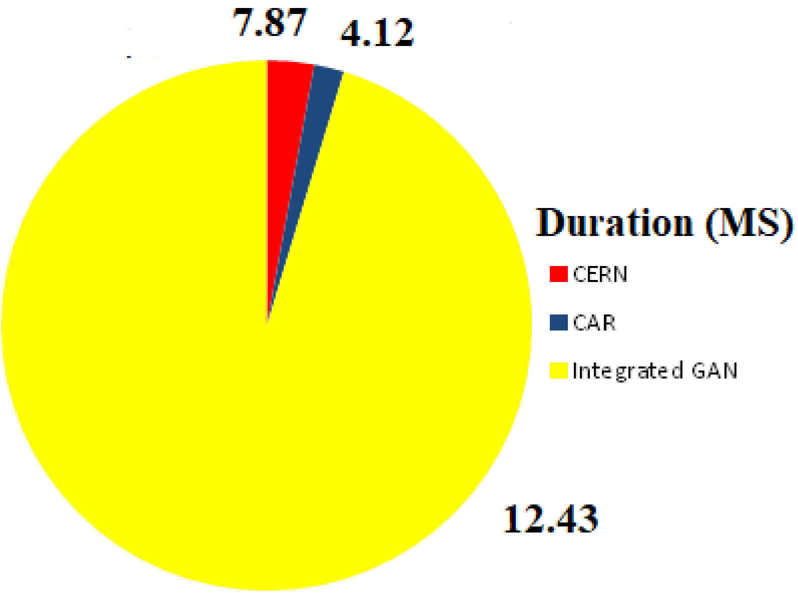

**Figure 6: Shows the Duration to process the Image**

## Conclusion

The proposed approach in this paper mainly focused on detecting and recognition various collective activities from the given datasets. The proposed model also focused on effective training on various human activities by using single and group. This model detects every person's activity based on the input video. The overall framework is trained in a weakly supervised manner, which represents the bounding boxes based on the collective activities which need labels. The proposed model achieved a huge accuracy of 87.98%, Sensitivity is 88.54%, Specificity is 88.32, F1-Score is 61.32 and duration is 4.12 sec compared with existing models.

## References

[1] [1] Behera A., Chapman M., Cohn G and Hogg. “Egocentric Activity Recognition using Histograms of Oriented Pairwise Relations”. In International Conference on Computer Vision Theory and Applications, pp. 22-30. (2012).

[2] [2] Fathi, A., Farhadi, A., and Rehg, J. M. “Understanding egocentric activities”. In International Conference on Computer Vision (ICCV), pp. 407– 414,(2011).

[3] [3] Bay, H., Tuytelaars, T., and Gool, L. V. SURF: “Speeded up robust features”. In European Conference on Computer Vision (ECCV), pp. 404–417, (2006).

[4] [4] Behera, A., Cohn, A. G., and Hogg, D. C. “Workflow activity monitoring using dynamics of pair-wise qualitative spatial relations”. In International conference on MultiMediaModelling (MMM), pp. 196–209. (2012).

[5] [5] Matikainen, P., Hebert, M., and Sukthankar, R.“Representing pairwise spatial and temporal relations for action recognition”. In European Conference on Computer Vision (ECCV), pp. 508–521. (2010).

[6] [6] Shechtman, E. and Irani, “Matching local self similarities across images and videos”. In Conference on Computer Vision and Pattern Recognition (CVPR),(2007).

[7] [7] Liu, W., Li, S., and Renz, J. “Combining rcc-8 with qualitative direction calculi: Algorithms and complexity”. In Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI), pp. 854–859. (2009).

[8] [8] Carneiro, G. and Lothey, D. “Sparse flexible models of local features”. In European Conference on Computer Vision (ECCV), pp. 29–43, (2006).

[9] [9] Lothey, D.G. “Distinctive image features from scale invariant keypoints”. International Journal of Computer Vision(IJCV), pp.91–110. (2004).

[10] [10] Fathi, A., Ren, X., and Rehg, J. M. “Learning to recognize objects in egocentric activities”. In Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3281–3288. (2011).

[11] [11] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, “Deep learning for human activity recognition: A resource efficient implementation on low-power devices,” in Wearable and Implantable Body Sensor Networks (BSN), 2016 IEEE 13th International Conference on. IEEE, 2016, pp. 71–76.

[12] [12] T. Plotz, N. Y. Hammerla, and P. Olivier, “Feature learning for ac- ¨ tivity recognition in ubiquitous computing,” in IJCAI ProceedingsInternational Joint Conference on Artificial Intelligence, vol. 22, no. 1,2011, p. 1729.

[13] J.Nageswara Rao, M.Ramesh,” A Review on Data Mining & Big Data, Machine Learning Techniques”, International Journal of Recent Technology and Engineering (IJRTE) , ISSN: 2277-3878, Volume-7 Issue-6S2, April 2019.

[14] D. Veeraiah and J. N.Rao, "An Efficient Data Duplication System based on

Received 05 May 2021; Accepted 01 June 2021.

Hadoop Distributed File System,"2020 International Conference on Inventive Computation Technologies (ICICT), 2020, pp. 197-200, doi: 10.1109/ICICT48043.2020.9112567.

[15] J. N. Rao, A. C. Singh, “A novel encryption system using layered cellular automata,” International Journal of Engineering Research and Applications, vol. 2, no. 6, pp. 912–917, 2012.

[16] J.Nageswara Rao, Bhupal Naik, G Sai Lakshmi, V Ramakrishna Sajja, D Venkatesulu, “Driver’s Seat Belt Detection Using CNN” Turkish Journal of Computer and Mathematics Education Vol.12 No.5 (2021), 776-785 3.

[17] J.N Rao, Dr.Rambabu Busi, Dr. G Rajendra Kumar, U. Surya Kameswari, “ Content image Retrieval Based on using open Computer Vision and Deep Learning Techniques “International Journal of Advanced Science and Technology,Volume29Issue03Pages5926 - 593 92020)