International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

A Survey and High-Level Design on Human Activity Recognition

Abhishikat Kumar Soni[1], Dhruv Agrawal[2], Md. Ahmed Ali[3], Dr. B. G. Prasad[4]

[1, 2, 3, 4]Department of CSE, B.M.S. College ofEngineering, Bangalore, India

Abstract: ***Human Activity Recognition is one of the active research areas in Computer Science. The domain is being recognized as it has a vast scope and immense potential. The resultant applications are being deployed in various contexts like security surveillance, healthcare, human-computer interaction, Traffic Control, and many more, will help minimize human errors and associated costs and increase the efficiency of the desired work. This work surveyed recent eleven research papers and the relevant topics in the Human Activity Recognition (HAR) domain. Different applications, such as detection of single or multiple persons, classification of simple activities, anomalous activities, tracking, etc., with proposed related technologies, are covered. Based on the review and the conclusion got, a feasible vision-based model to detect, classify and recognize human activities or actions has been proposed. Techniques like Haar feature-based technique [10] [11] for the detection of human beings and feature extraction, YOLO [5], CNN (Resnet) [1] [2] [4] The classification and recognition of the defined activities have been proposed to be used. The design includes sub-systems such as human detection and feature extraction, pre-processing of the frames, classification of actions using Deep Neural Network, and raising warnings. The system has been proposed to work within the limited computational capabilities to determine anomalous activities.***

Keywords: Human Activity recognition, Vision Based Hu- man Activity Recognition

1. Introduction

Human activity recognition plays a significant role in human-to-human interaction and interpersonal relations. Because it provides information about the identity of a person, their personality, and psychological state, it is difficult to extract. The human ability to recognize another person's activities is one of the main subjects of study in the scientific areas of computer vision and machine learning. Many applica- tions, including video surveillance systems, human-computer interaction, and robotics for human behavior characterization, require a robust multiple activity recognition system. Among various classification techniques two main questions arise: "What action?" (i.e., the recognition problem) and "Where in the video?" (i.e., the localization problem). When attempting to recognize human activities, one must determine the kinetic states of a person, so that the computer can efficiently rec- ognize this activity. Human activities, such as "walking" and "running," arise very naturally in daily life and are relatively easy to recognize. On the other hand, more complex activities, such as "peeling an apple," are more difficult to identify.Complex activities may be decomposed into other simpler activities, which are generally easier to recognize. Usually, the detection of objects in a scene may help to better understand human activities as it may provide useful information aboutthe ongoing event.

1.1 Motivation

There is plenty of scope in the Human Action Recognition domain. One can characterize actions on the basis of normal and abnormal behaviour. Different kinds of activities that a human performs during his day-to-day chores can be detected and classified. In particular, activities related to a closed environment can be examined as an application to some defined task. For example, activities like running or fall canbe defined as abnormal with respect to the decorum of a place and the system can be applied to detect such activities and raise warnings at appropriate times.

1.2 Objective

Study of various research papers in the Human Activity Recognition field and to consequently come up with a feasible system to detect and recognize human actions has been the objective of this work. Initially, a survey on relevant research papers has been conducted to study various technologies and tools currently employed in the HAR domain. Consequently, some of the best algorithms and techniques that could be suited to the need and limitations have been found. Then, a feasible system with high-level design employing adequate techniques to fulfill the desired task of recognizing and classifying activ- ities has been proposed.

1.3 Scope

The scope of HAR Systems is vast and varied. Video surveillance and human actions detection and recognition is becoming a very popular and most sought after discipline. Videos and images are generated everyday and in a huge amount. Using the technologies, many researchers are dedicating their efforts to make this huge data useful and trying to eliminate human efforts as far as possible by deploying intelligence in the new systems. Various applications like monitoring, controlling, warning systems are being proposed. General Procedure of a simple HAR system is to first detect the human actions, then classify and lastly recognize them as defined internally. Major drawback of the domain is that itis still new and very complex. High GPU requirements and other computational capabilities along with present hardware infrastructure pose major economic problems. Despite these challenges, there is an immense scope in the field. Throughits applications many human-based efforts will get automated and the efficiency of the desired work will increase with decrease in the costs. Simple example would be an automated monitor system in the lifts to detect the owners of dogs which dirty the

premises. Also, motion based surveillance systems in public places will be extremely useful where the cameras will automatically start monitoring for human presence in the premises.

1.4 Existing System

The existing systems use methods like OpenCV and Tensorflow or inefficient combinations of existing techniques to detect activities for which high processing time is incurred. The system presented in S. Adarsh et al. [5] uses two neural networks, YOLO for human detection and ResNet-34 for activity classification, employing the use of input frames at both levels. Abhay Gupta et al. [8] mentions that the use of wearable device-based and smartphone based techniques are becoming popular. The inbuilt sensors like gyroscope are used to collect the body gestures information which is then processed with some AEs to get the prediction. Gatt et al. [9] uses pre-trained post estimation models such as OpenPose and PoseNet along with LSTM architecture to classify actions into normal and abnormal actions. They use 2D skeletal data of normal cameras as input feed and extract the body keypoints and accordingly reconstruct to analyze error which leads to classification. Various handcrafted technologies such as Hidden Markov Models (HMM) and Support Vector Ma- chines (SVM) were proposed in earlier studies. Recently most researchers use autoencoders (AEs), CNNs, LSTM or their combinations. Atikuzzaman et al. [10] demonstrated the use of Haar feature-based classifier to detect human poses and CNN to recognize different classes of activities. Normal CCTV videos and camera images were fed to the system without the need of any specialized depth cameras and wearable device sensors.

1.5 Proposed System

We have developed a proposed system according to the survey conducted on below recent research papers. The Kinetics dataset (the dataset we will use to train our human activity recognition model) can be used to train the model as it will reduce the pre-processing time [1]. Also, training the Resnet-34 model on the kinetic dataset may overcome overfitting, as suggested by Shikha et al. [2]. In addition, to the kinetic dataset, we proposed to include the UCF-50 Action Recognition dataset so that more activities can be classified. The input feed will then be processed with a Haar feature-based classifier for human detection and feature extraction. After that, output frames can be pre-processed by Ten pre-processing systems for filtering the frames and getting positive frames with an adequate human presence on which the system could be trained. The activities then need to be recognized using the YOLO technique. Later on, we proposed to create an introductory video classification system with Keras to classify the activities into normal and abnormal ones. These classifications will be internally defined according to the desired need. For example- an everyday activity like running can be classified as abnormal for achieving the desired task. Additionally, we are planning to create a normal classifier, then implement an average moving technique, and finally make use of Single Frame CNN technique for all detection and classification activities. Furthermore, at last, A warning or alert raising output to notify the users that some abnormal activity has happened and been detected by the system. For example- the color of the box containing humans may change to red color.

2. Literature Survey

2.1 Maintaining the Integrity of the Specifications

Akansha A. et al. [1] present an introduction regarding the relevance of combining object detection and tracking technologies in computer vision. It describes the technology used for object detection, which uses optical flow, back difference, and frame difference methods that, based on the continuous motion of the objects, divide them into grids, classify the boxes and construct the target item. The authors suggest using the Resnet-34 model for the image classification of convolutional neural networks. Deep neural networks are utilized, which assists in the extraction of more relevant characteristics. Classification methods are then used on these characteristics. A kinetic dataset that comprises 400 types of human activity is utilized for training the model. For additional study, the authors recommend that utilization of a dataset with more than 400 actions might be developed in order to raise the degree of accuracy and make the system more adaptable. It is noted that if there is a deep hierarchy of activities like yoga with multiple poses, a dance with diverse forms, and many other similar activities, then it might considerably aid in higher performance. The authors cover issues and limits, including recognition of complicated and simultaneous actions. These actions become confused and difficult to distinguish. Sensor-based technologies also confront various obstacles like the implantation of devices on different regions of human bodies to assess activities directly. It becomes difficult for consumers to wear sensors implanted in their watches, garments, wristbands, etc. In an intelligent house, sensors must be put in every door and equipment. Installation and maintenance of such a massive network are rather laborious.

The study offered by Shikha et al. [2] presents a step-by- step process to create a human activity recognizer. The overall architecture of the Resnet model is given first, along with a description of its process, followed by the technique and outcomes of the implementation. The trained model offers an accuracy of 79 The authors propose using a Convolutional neural network (CNN) with spatiotemporal three-dimensional (3D) kernels such as Resnet-34 that is trained using Kinetics data set, which includes 400 classes. For future study, the authors recommend a possible use of a dataset containing more than 400 activities to make the system more adaptable, for example, having a more comprehensive dataset that splits the different yoga asanas into different labels. With this, it is noted that increasing the number of samples for an activity in the dataset will enhance the system's performance. The authors address the issues and constraints with the accuracy of complex activities that were significantly lowered for activities like cooking, yoga, etc. since there were numerous methods of conducting such activities.

Pankaj B. et al. [3] targeted building a cost-effective and speedier Human Activity Recognition System which can

analyze both video and picture to detect the activity being conducted in it, hence assisting the end user in many applications like monitoring, aiding purpose, etc. The system mentioned will be not only cost-effective but also a utility-based system that can be implemented in many applications that will save time and help in many tasks that need identification process and save a lot of time with excellent accuracy. The model exhibits strong results on video streams while performing reasonably on picture data. The authors propose to develop the solution utilizing neural network architecture. This architecture uses resnet-34 pre-trained on kinetics dataset of 400 classes depicting activities of humans in their everyday life for processing the videos and is further refined using the transfer learningon more concentrated activities while using the caption generation technique on the images. For future study, the author believes that video recognition code may be better adjusted using transfer learning and that much bigger datasets can be employed to boost the model's accuracy further. Moreover, online and mobile apps may be constructed which can contact these python scripts via an API request to give activity recognition on users mobile, and can also benefit the old and blind people. The authors address issues and limits, including the identification of complicated as well as simultaneous actions, which makes it tougher to discern between two or more activities of close locations.

Akash K. et al. [4] worked on constructing a Human Activity Recognition System to recognize human activity being conducted on the input stream, such as a pre-recorded video or a live video input. One of the critical workings of this system is to recognize the activities that humans undertakeand consequently tag them. The trained model can identify the activities with an accuracy of 70 to 90 percent. The authors presented a model based on Python 3, Keras, OpenCV, ResNet, and TensorFlow. The biggest reason for this deviceis to system the enter video flow for human identification and, similarly, process the character frames of the enter video to expect which hobby is being carried out with the help of employing the human. After the prediction is created that is as correct as 94 percent, the frames are captioned, and the ultimate result is presented to the output. For future study,the authors recommend that a more excellent and effective program be worked upon to understand and identify sporting activities in long-duration recordings. The authors discuss the problems and drawbacks of comparison amongst human sports because of the terrific form of strategies carried out to symbolize likeness and the dependence that the outcomes give from the used dataset, hence stressing the need and importance of a dataset with varied captures.

Adarsh S. et al. [5] focussed on a surveillance system that utilizes Human Activity Recognition algorithms to identify whether the objective human is an ordinary individual efficiently or a suspicious one is described and implemented. The accuracy gained by the suggested approach is determined to be 82 percent. The authors recommended ideally usingthe input frames from the film captured in the security camera, further which is processed by a YOLO model pre-trained on the COCO dataset, where the persons presentin the footage are spotted. After which, the frames are transferred to the Residual network (ResNet-34) model. The ResNet– 34 is trained with the UT-Interaction dataset to categorize the actions in the input frames. Finally, afterthe categorization, the conduct is tagged on the screen accordingly. In the future, the detection algorithms might be further adjusted to use fewer frames to identify and categorize the activity. It can be developed to identify more complicated behaviors in diverse and more complex contexts. The authors address issues in identifying complicated and few simultaneous actions, making it harder to discern between two or more activities of exact locations. This limits the ability of the model to perform efficiently in the above-said scenarios.

Introduction with the definition of Artificial Intelligenceand Machine Learning is given by Sumita Das et al. [6] inher paper. The importance of AI/ML in today's world is explored through investigating numerous applications of AI, and machine learning as a branch of AI has been under focus. Types of Machine Learning Algorithms, i.e., Supervised learning, Unsupervised Learning, Reinforcement Learning, and Recommender Systems with various applications, have been emphasized to have good insights into the technology deployed and the domain in which they are applied. Supervised learning comprises training and testing, anticipated output v/s computing output examination, and analysis. Pattern Recognition, Speech Recognition, Improvising Operating Systems, etc., are some of the uses of this technology. Unsupervised learning is frequently referred to as learning on its own. An input data stream is delivered, and categorization is done for distinct clusters based on specific internal classification rules. It does not require considerable training. DNA Classification, Market Segmentation, Speech Activity detection, etc., are some applications connected. Reinforcement Learning incorporates training methods based on rewarding desired behaviors and punishing unwanted ones. An agent must execute actions that optimize performance or give the most reward in the set environment. Traffic forecasts, Computer Games, and Stock Market Analysis represent some of the uses. Recommender System teaches a machine a basic principle- "learning to recommend." Applications are Advertisements, Sentiment/Opinion Analysis, and self-customizing programming. With the quantity of data that is accessible and is created every day, the use, storage, and management of data is a worry and difficulty. Big Data and Data Warehousing technologies have become necessary for storing and administering such enormous data. The utilization of data depends on the shoulders of AI/ML researchers who supply adequate approaches to apply analytics and use the meaningless data. Nevertheless, Data sets need to be regularly updated and maintained up-to-date as learning is a continuous process. Technologies must be developed to train a computer with massive datasets in less processing time. Hardware expenses need to be cut. Apart from these restrictions, AI/ML has an enormous scope. Humans constantly desire their lives to grow more pleasant, and in the present, it is feasible to make machines more intelligent. Machine Learning has been substantially effective in various domains such as data mining, OCR, statistics, computer vision, mathematical optimization, etc.

G. Sreenu et al. [7] have studied deep learning techniques used for human activity recognition, notably concerning crowd analysis. The study presented a deep-rooted survey that started with object identification, action recognition, crowd analysis, and violence detection in a crowded context. Most studies evaluated in the paper are based on deep learning approaches frequently employing CNN, Auto-encoders, and their combination. The emphasis has been on applying deep learning algorithms in determining the accurate count, participating participants, and the occurring action in a vast crowd. The author emphasized that video surveillance data has much social relevance in today's world among the various data sources which contribute to terabytes of big data. The same could be used for various purposes like monitoring, surveillance, control, alarm, and security, and the so-called senseless data may soon become most valuable by employing some intelligence. Handling this extensive data promotes the application of technologies like Big Data, Data Warehousing, Data Mining, and Data Analytics. So, logically these technologies also have to be regarded. Some deep learning techniques examined are YOLO, LSTM, BPTT, AMDN, Stacked denoising, and autoencoders. Combinations of several of these strategies have also been discussed. Real-time processing has been a fundamental problem that still needs to be investigated extensively in this sector. Crowd analysis still requires much work as it has been the most challenging component. All forms of activities, behavior, and movement need to be identified. Moreover, crowd size has been vast and changeable in actual-world circumstances. So, the report indicates that although video surveillance has highly essential and different uses, many efficient technologies need to be investigated, notably crowd analysis.

Three popular methods of recognizing activity in Abhay Gupta et al. [8], namely vision-based (using pose estimation), wearable devices, and smartphone sensors, have been discussed. The task is identifying and recognizing actions or activities that a person performs. The wearable device-based technique uses some sensing devices to be mounted on the subject to collect sensor data. The smartphone-based technique uses smartphone sensors such as a gyroscope and accelerometer. In vision-based techniques, estimation is done using an individual's pose, that is, by analyzing key body points through neural networks and predicting the abnormality in their pose. Some related works in the HAR field have also been reviewed and put forth by the author. Various effective technologies used by the researchers include pose-based HAR, CNN, Kinect Sensor, Hidden Markov Models (HMMS), PoseNet, and OpenPose for pose estimation model and LSTM. Smartphone sensors, gyroscopes, and accelerometers have also been put forth for human activity detection. In the end, a table comparing all the three techniques in the HAR domain has been shown to have a better insight into their applicability and resultant accuracy. It was put forward that smartphone-based and wearable device-based techniques are more popular than vision-based techniques. Despite this, the comparison revealed that the vision-based approach is more versatile as it manages to classify and recognize more actions than both other techniques and is more accurate. However, Human Activity Recognition (HAR) remains a complex task due to unresolvable issues such as sensor movement, sensor placement, background clustering, and the inherent variability of how different people perform activities. The vision-based approach isthe least popular among the three techniques. Detecting and extracting people from image sequences requires sophisticated machines. Furthermore, cameras are required for the purpose, which has associated privacy issues. Also, cameras need to be bought and mounted on some support, increasing the cost. Despite limitations, the Vision-based technique can become an excellent choice for HAR. With the advancements in technology and the availability of machines with high computational power, a vision-based technique has a more potential scope in the future. The technique will help to increase the detection efficiency of some undetermined actionsand will also help to include and detect more versatile actions.

Thomas Gatt et al. [9] proposed an approach based on collecting and examining the coordinates of body movement for human activity detection. An automated camera-based system able to detect irregular human behavior is put forward by the author, basically to detect the fall activity that lies under the definition of abnormal activity. "Abnormal activities are those rare events that deviate from normality" [9]. For example, using the joint positions of the hip and shoulder centers, the person's torso angle is calculated when such an angle exceeds a threshold. Then, it can be analyzed for a fall. The author propounds that deep-learning-based techniques, especially CNNs, have been employed majorly due to various complexities involved in video analysis. PoseNet and OpenPose, pre-trained pose estimation models, have been used to detect the person in the frame and extract the body key points. The extracted data is then fed to auto-encoders (LSTM, CNN) to learn a general representationof the expected behavior, consequently alerting to abnormal behavior when detected after classification. Pose estimated data can be informative enough to understand and classify human actions as normal and abnormal has been suggested by the study. Still, some challenges need to be acknowledged. HAR is still a complex task and needs innovation, and more improved methods as most of the methods involved are hand-crafted methods like HMM and SVM and require domain expertise in areas such as statistics, mechanics, etc. A single human being was tested in the experiment showing that crowd analysis and capturing several humans and classifying their activities is still a challenge. As always, computational limitations and the cost of infrastructure is an issues. As proposed by the author, some future work is the extension of the model to use 3D Pose Estimation in addition to the2D model used here. Also, more actions could be added and defined as anomalous for detection. They also suggest bringing down the computational requirement by using 2D pose data on a single board computer (such as Raspberry Pie) with a system-on-module featuring a GPU or a TPU.

Atikuzzaman et al. [10], has focussed on Haar feature based classifier for detecting human presence and their poses along with a CNN classifier to recognize and classify the human actions. CCTV footage and images from cameras were employed as input data streams. The use of CNN has

been an inevitable part of the process. A number of modules such as object detection, segmentation and recognition are crucial as CNN works in steps and as all have to work sequentially to have the correct output. The system is put forth with the aim to detect human activity and to recognize different classes of body movement from videos. Walking, running, standing, sitting and lying are the 5 basic activities that have been considered for recognition. The proposed system is also trained on their self-collected dataset composed of 5648 images. Dataset and system design are the two suggested approaches. Further, System Design is divided into three categories: human detection/localization, segmentation and video frame recognition. Images were turned to gray scale to eliminate the need of using a GPU. An efficacious detection accuracy of 99.86approximately 22 frames per second after 20 epochs is asserted by the author to be achieved through the approach. Some limitations encountered were the lackof proper computing infrastructure and heavy costs incurred for obtaining the same. Due to the same, only 5 defined categories of actions were trained and tested. Also, due to lack of standard dataset as the dataset was made by their own, the system couldn't be tested properly with distinct weather and light conditions. Also, the paper is about just the detection and recognition of some defined normal activities. There is no information regarding anomalous or abnormal activities. Future scope as proposed will test more range of activities by using more standard and large datasets. Also, a scope for developing systems which could detect human behavior is put forth. On the same, behavior then can be classified into normal and abnormal.

The detection and tracking of human beings from video footage is the focus of Visakha K et al. [11]. The video feed can be pre-processed or live. The system works with crowds and tracks every human detected using a sampling-resampling algorithm. The entire process is divided into two main modules- Human Detection and Human Tracking. A video surveillance system consists of three phases: moving object recognition, tracking, and decision making. Detectionis made possible by using the functions of the Haar Classifier. The proposed classifier consists of multiple simpler classifiersapplied to a particular region of interest in a series of stages until the candidate is rejected or passes all the stages at some stage. The proposed system can also detect anomalous activ- ities using detection and tracking mechanisms, for example, detecting abnormal activities such as human presence in a shopping complex after regular working hours and raising warnings against the activity. Haar classifier is a cascade system that performs the tasks in steps. So, Haar can be used with other technologies to achieve better efficiency and calculation speed with an adequate computational setup. The work can be extended for control application areas like traffic control and monitoring, where automatic photos could be taken on violation of traffic laws, urban surveillance, home security, etc. The system can be extended by detecting anomalous activities like falls, violence, and running to trigger warnings in an emergency.

3. Requirement Analysis and Specification

As per the survey conducted in various research papers, we found out that the below functional and non-functional requirements are the minimum requirements a user has to fulfill to get the accurate result for human activity recognition.

A. Functional requirements

Functional requirements are the features that will be demonstrated by the system during its course of use. These requirements decide the usability of a system. If the requirements are proposed clearly, the system will be robust and very efficient. Dataset Training: Attributes of activity details will be storedin the form of the dataset, and through various proposed algorithms, the machine learning model will be trained. The system should take the input, analyze the feed, and get trained. Derivation of frames: The system should derive frames from the video input, and time-dependent attributes should be maintained. The frames should be derived in the proper order. Analysis of Frames: The derived frames should be analyzed. The presence of humans (positive frames) and the absence of humans (negative frames) should be examined. Accordingly, the positive frames with better quality of human presence and resolutions should be fed forward.

Detection of activities: The human activities should be detested by the system after getting the positive frames. The detection should depend on the internal definition of human activities.

Classification of activities: The system should, after detecting the activities should, classify the activities according to the internal definition in terms of normal and abnormal activities. The system should be equipped to intercept abnormal activities as they fall against some threshold, and everyday activities will continue to be allowed.

Recognition of Activities: The system should recognize different activities proposed per internal definitions. Abnor- mal activities should be recognized as per their types. Output Warning: The system at the end should give a comprehensible output. A warning signal should be given to let the user recognize that something abnormal has happened.

B. Non-Functional requirements

Nonfunctional requirements define how a system should behave and place practical limits on it. The system's quality attributes are another name for this type of requirement. Performance, security, usability, and compatibility are not features of the system; rather, they are necessary characteristics. We can't create a specific line of code to execute them because they are "growing" properties that arise from the entire arrangement. The specification describes any attributes that the customer requires. Only the requirements that are relevant to our project should be included.

Dependability: In order to provide the functionalities, the structure must be reliable and sturdy. When a consumer has revealed a couple of improvements, the structure must make the actions clear. The Programmer's advancements mustinclude Project pioneering as well as Test designer.

Affordability: The monitoring and maintenance of the system should be fundamental and focused in its approach. There should not be an excessive number of jobs running on sev- eral machines, making it difficult to monitor if the jobs are operating smoothly.

Execution: Throughout the process, multiple representatives will use the framework. Because the system will be run on a single web server with a single database server hidden from view, execution will change.

Reliability: It is very important that the system be reliable. It should not frequently go into halt state, and work properly. The users should be able to rely on the system for its usability and should be worry-free regarding its functioning. The systemshould demonstrate the user's trust.

Manageability: The system should easily be manageable. The user should learn the system in less time and use it conve- niently. The managing and operating costs and labour should not increase significantly over the time and the system should be easily deployed for the intended use.

4. Design

Initially, input has been taken from a camera (can be DSLR, CCTV etc.) after that object recognition method has been imposed on it. The video has been split into multiple frames and afterwards Image processing has been done. Image processing is done with the help of Keras. Now, the Activity Recognition System processes the data to classify the activity accordingly and at last if something goes wrong like the activity is Abnormal or anomalous then a trigger alert is generated which warns the user for the same.

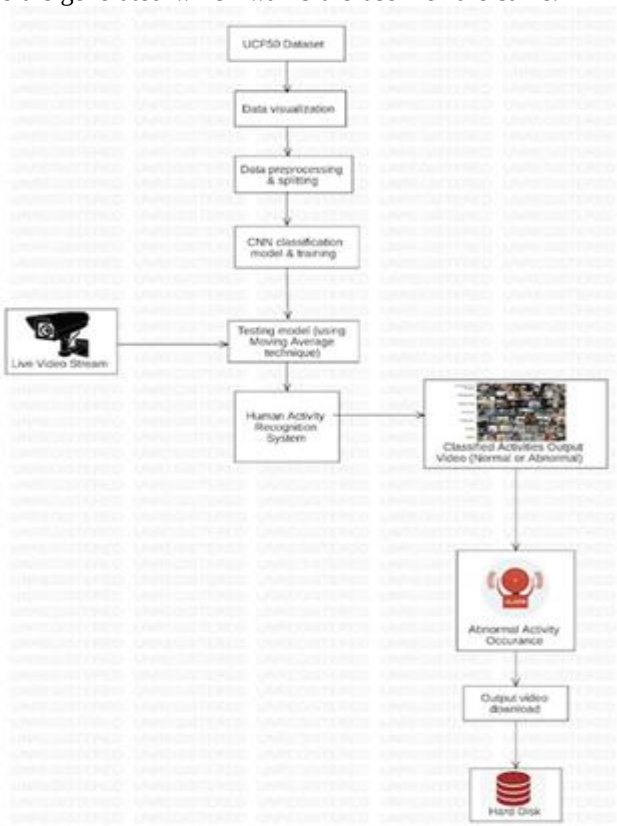

Figure 1: High Level Design for Activity Recognition

A. Abstract specification of Subsystems

Hybrid technologies will be used to construct the machine learning model. As a result, the model may be used to train any dataset. Object-oriented interactions of various model components are depicted

Dataset acquirement: The dataset is acquired if already available and if the dataset is not available then it is scraped from various platforms that house reviews regarding insurance companies using web-scraping. The dataset is composed of reviews given on different insurance companies.

Pre-processing: This step is one of the most important steps of all. This is because the better the pre- processing the higher will be the model performance and accuracy. Thus to make sure relevant pre-processing steps are taken into consideration we make a checklist of different pre-processing steps. They are as follows: Video of larger size has to be compressed in order to process it. Unnecessary video frames can be cut down to decrease the pre-processing time.

Splitting of the Dataset: Dataset is split into training and testing parts. If required a validation set is also employed.The training dataset is the major part of the dataset on which the model will be trained. The testing dataset is the minor part of the dataset on which the dataset will be evaluated using different metrics. If a validation set is employed it will be for hyperparameter optimization.

B. Methodology

As per the literature survey, we have proposed a method:
Step 1: Download and Extract the Dataset – We can use the Kinetic Dataset as it has plenty of activities. As per theconvenience, data can be altered and reduced. The datasetWe are suggesting the UCF50 – Action Recognition Dataset.The dataset has 400 video clips per class (downloaded viaYouTube) and 300,000 videos.
Step 2: Visualize the Information with its Names and labels.
Step 3: Read and Preprocess the Dataset.
Step 4:Part the Information into a Prepare and Test Set.
Step5: Develop the Model.
Step 6: Compile and Train the Model.
Step 7: Plot Model's Misfortune and Exactness Bends Step
Step 8: Make Predictions with the Model.
Step 9: Utilizing Single-Frame CNN Strategy

The Kinetics dataset is also sufficiently large, and therefore, these architectures should be able to perform video clas- sification by changing the input volume shape to include spatiotemporal information and utilizing 3D kernels inside of the architecture. By modifying both the input volume shape and the kernel shape, we are expecting to obtain great accuracyon the Kinetics test set, and a similar accuracy on the UCF-101test set and on the HMDB-51 test set.

5. Conclusion

With the enormous amount of video and image data col- lected today and the presence of advanced technologies like Big Data, Data Mining, Data Warehousing, Machine Learning, Deep Learning, Analytics, and Artificial

Intelligence, the task of incorporating intelligence into machines for better efficiency and elimination of human labor is getting hugely recognized. It is becoming a need of the hour. Several technologies and techniques have been proposed and developed for the same. From the three most popular methods, namely, Vision-based, Smartphone-based, and wearable device based, it was found that smartphone-based and wearable device-based methods are becoming popular. Although the vision-based technique is not so popular due to its infrastructure requirements, studies suggest that this technique is the most efficient one and will evolve significantly in the future. The technologies discussed correspond to the system's intended use and applicability. Pose-based estimation using PoseNet and OpenPose is suggested for single human abnormal activity detection. Haar feature-based classifier is suggested for detecting everyday human actions like running, walking, laying, and many more. However, using CNN is inevitable and is the most crucial technique that needs to be used. Also, various combinationsof autoencoders, LSTM, CNN, Pose estimation, and Haar classifier were put forward to achieve better efficiency. Resnet- 34, a 34-layer convolutional neural network, can be utilized as a state-of-the-art image classification model. The Kinetics dataset, which is a large-scale, high-quality dataset for human action recognition through videos, could be utilized to train the Resnet-34 model. However, various challenges still needto be tackled. Computational capability and adequate infras- tructure is the most significant limitation. However, with the rapid advancements being made every day in electrical and hardware technology, this limitation could be tackled with the availability of cheap capable hardware shortly. Another limitation is the captured human posture in the image. Posture must be correctly captured to filter and recognize the activity. Height, weight, depth, and positioning of an object, along with the lighting conditions, matter for an effective detection. Management of datasets is another challenging task. Datasets that have been used have to be updated regularly to acheiving great accuracy in the long run.

References

[1] A. Akansha, S. Anisha, K. Kritika and R. Raju, "Human Activity Recognition using Resnet-34 Model", International Journal of Recent Technology and Engineering (IJRTE) (India), vol. 10, issue 1, May 2021,ISSN. 2277-3878.

[2] Shikha, K. Rohan, A. Shivam, J. Shrey, "Human Activity Recognition", International Journal of Innovative Technology and Exploring Engineer- ing (IJITEE) (India), vol. 9, issue 7, May 2020, ISSN. 2278-3075.

[3] B. Pankaj, K. Harpreet, G. Akarshit, S. Jaskaran, "Human Activity Recognition System", Oriental Journal ofComputer Science and Tech- nology (India), vol. 13, issue 2-3, pg. 91-96, October 2020, ISSN. 0974-6471.

[4] K. Akash, S. Varshini, T. Puneet, "Human Activity Recognition System", International Journal ofAdvance Research, Ideas and Innovations in Technology (IJARIIT) (India), vol. 7, issue 3, May 2021, ISSN. 2454- 132X.

[5] S. Adarsh, B.S. Vidhyasagar, K. Giridhar, J. Arunnehru, S. Poorvaja, "Suspicious Activity Detection And Tracking In Surveillance Videos", Journal of Emerging Technologies and Innovative Research (JETIR), vol. 7, issue 5, May 2020, ISSN. 2349-5162.

[6] Sumit Das, Aritra Dey, Akash Pal, Nabamita Roy, "Applications of Artificial Intelligence in Machine Learning: Review and Prospect", International Journal of Computer Applications, Volume 115 - No. 9, April 2015.

[7] G. Sreenu, M.A. Saleem Durai, "Intelligent video surveillance: a reviewthrough deep learning techniques for crowd analysis", Journal ofBig Data 6, Article Number: 48 (2019).

[8] Abhay Gupta, Kuldeep Gupta, Kshama Gupta, Kapil Gupta, "A Sur- vey on Human Activity Recognition and Classification", International Conference on Communication and Signal Processing (ICCSP), DOC: 28-30 July 2020, Chennai, India.

[9] Thomas Gatt, Dylan Seychell, Alexiei Dingli, "Detecting human ab- normal behavior through a video generated model", International Sym- posium on Image and Signal Processing and Analysis (ISPA), ISSN: 1849-2266, DOC: 23-25 Sept. 2019, Dubrovnik, Croatia.

[10] Md. Atikuzzaman, Tarafder Razibur Rahman, Eashita Wazed, Md. Parvez Hossain, Md. Zahidul Islam, "Human Activity Recogition System from Different Poses with CNN", International Conference on Sustain- able Technologies for Industry (STI), DOC: 19-20 Dec. 2020, Dhaka, Bangaladesh.

[11] Visakha K, Sidharth S Prakash, "Detection and Tracking of Human Beings in a Video using Haar Classifier", International Conference on Inventive Research in Computing Applications (ICIRCA), DOC: 11-12 July 2018, Coimbatore, India.