Volume 9, Issue 4, April – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1530

Human Activity Recognition using Machine Learning

Omkar Mandave[1]; Abhishek Phad[2]; Sameer Patekar[3]; Dr. Nandkishor Karlekar[4]
Mahatma Gandhi Mission's College of Engineering and Technology, Navi Mumbai, Maharashtra

Abstract:- Working information and classification is one of the most important problems in computer science. Recognizing and identifying actions or tasks performed by humans is the main goal of intelligent video systems. Human activity is used in many applications, from human-machine interaction to surveillance, security and healthcare. Despite continuous efforts, working knowledge in a limitless field is still a difficult task and faces many challenges. In this article, we focus on some of the current research articles on various cognitive functions. This project includes three popular methods to define projects: vision-based (using estimates), practical devices, and smartphones. We will also discuss some advantages and disadvantages of the above methods and give a brief comparison of their accuracy. The results will also show how the visualization method has become a popular method for HAR research today.

Keywords:- *Artificial Intelligence (AI), Human Activity Recognition (HAR), Computer Vision, Machine Learning.*

I. INTRODUCTION

Skeleton features are widely used in human action recognition and human-computer interaction [1]. This technology involves detecting and tracking key points of a human skeleton from images or videos, typically utilizing depth cameras, sensors, and other equipment to capture movement trajectories and analyse them through computer vision and machine learning techniques. The applications of skeleton behaviour recognition span across various domains such as games, virtual reality, healthcare, and security [1].

However, recognizing actions from skeleton data poses several challenges due to factors such as multiple characters in videos, varying perspectives, and interactions between characters [1]. Early methods relied on hand-designed feature extraction and spatiotemporal modelling techniques, while recent advancements have been made in deep learning-based approaches, particularly focusing on skeleton key points or spatio-temporal feature analysis [1].

Deep learning methods have shown promise in improving action recognition accuracy, but traditional approaches still face challenges in adapting to different scenarios and effectively handling pose changes and high motion complexity [1].

For instance, the MS-G3D network addresses some of these challenges by automatically learning features from skeleton sequences and incorporating 3D convolution and attention mechanisms [1]. To further enhance the performance of action recognition systems, new methods like the 3D graph convolutional feature selection and dense pre-estimation (3D-GSD) method have been proposed [1]. This method leverages spatial and temporal attention mechanisms along with human prediction models to better capture local and global information of actions and analyse human poses comprehensively [1]. Meanwhile, in a separate line of research, the use of Convolutional Neural Networks (CNNs) for human behaviour recognition from different viewpoints has gained attention [2].

CNNs are utilized for tasks such as object detection, segmentation, and recognition in videos, aiming to identify various classes of body movement [2]. In parallel, the development of Human Activity Recognition (HAR) systems has been propelled by the demand for automation in various fields including healthcare, security, entertainment, and abnormal activity monitoring [3]. HAR systems rely on computer vision-based technologies, particularly deep learning and machine learning, to recognize human actions or abnormal behaviours [3]. These systems monitor human activities through visual monitoring or sensing technologies, categorizing actions into gestures, actions, interactions, and group activities based on involved body parts and movements [3]. The applications of HAR encompass diverse domains such as video processing, surveillance, education, and healthcare [3].

Similarly, action identification in videos remains a crucial task in computer vision and artificial intelligence, with applications in intelligent environments, security systems, and human-machine interfaces [4]. Deep learning methods, including convolutional and recurrent neural networks, have been employed to address challenges such as focal point recognition, lighting conditions, motion variations, and imbalanced datasets [4]. Proposed models integrate object identification, skeleton articulation, and 3D convolutional network approaches to enhance action recognition accuracy [4]. These models leverage deep learning architectures, including LSTM recurrent networks and 3D CNNs, along with innovative techniques such as feature extraction and object detection to classify activities in video sequences effectively [4].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1530

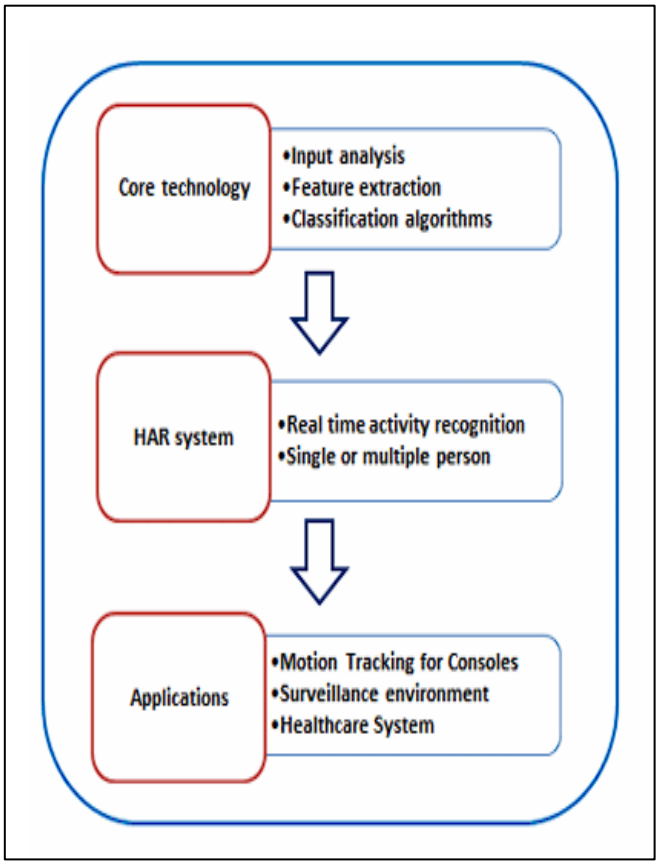

Fig 1 The Overview of General HAR System this is Basic Building Blocks of Almost Every System

➢ *Problem Statement & Objectives*

Human Activity Recognition (HAR) aims to develop algorithms and systems capable of identifying and classifying various human activities based on data collected from sensors or video streams. In fields such as healthcare, sports performance analysis, and human-computer interaction, accurate HAR systems can provide valuable insights into individual behaviors, health monitoring, and enhancing user experiences. However, challenges arise due to variations in sensor data quality, diverse activity patterns, and real-time processing requirements. Thus, the need arises for robust and efficient HAR techniques that can accurately classify human activities in real-time, considering the complexities of dynamic environments and diverse user contexts.

II. LITERATURE REVIEW

➢ *3D Graph Convolutional Feature Selection and Dense Pre-Estimation for Skeleton Action Recognition:*

The proposed method utilizes a 3D graph convolutional feature selection (3DSKNet) to adaptively learn important features in skeleton sequences. Additionally, a DensePose algorithm is introduced to detect complex key points of human body postures and optimize the accuracy and interpretability of action recognition. 3DSKNet focuses on key skeletal parts, improving the accuracy and robustness of bone recognition by selecting important features adaptively. DensePose provides more detailed pose and shape information, enhancing the accuracy of human motion analysis.

➢ *Human Activity Recognition System from Different Poses with CNN:*

Human Pose Detection: Utilizes HAAR Feature-based Classifier for detecting human poses in videos or images. Human Pose Segmentation: Segments human poses based on the output of the HAAR feature-based classifier. Activity Recognition: Uses Convolutional Neural Network (CNN) for recognizing human activities from segmented human poses. Human Pose Detection: Utilizes HAAR Feature-based Classifier trained on grayscale images to detect human poses accurately. Human Pose Segmentation: Segments human poses based on the coordinates obtained from the HAAR feature-based classifier and resizes the images to 64x64 resolution. Activity Recognition: Employs a CNN architecture, specifically VGG19, for recognizing human activities from the segmented human poses.

➢ *Effectiveness of Pre-Trained CNN Networks for Detecting Abnormal Activities in Online Exams:*

Detection of online exam cheating using deep learning, motion-based keyframe extraction, COVID-19 dataset collection, pre-trained models (YOLOv5, Inception_ ResNet_v2, DenseNet121, Inception-V3), and CNN fine-tuning for enhanced performance. Analysis of cheating activities during online exams using deep learning models.

Collection and preprocessing of a dataset containing videos of various cheating behaviors. Use of pre-trained and fine-tuned deep learning models for classification. Focus on model accuracy, loss, precision, recall, and F1-score for evaluation.

➢ *A Real-Time 3-Dimensional Object Detection Based Human Action Recognition Model:*

Data fusion technique to merge datasets. Utilization of 3D Convolutional Neural Network (3DCNN) with multiplicative Long Short-Term Memory (LSTM) for feature extraction. Object identification using finetuned YOLOv6. Skeleton articulation technique for human pose estimation. Combination of four different modules into a single neural network. Integration of data fusion, feature extraction, object detection, and skeleton articulation techniques. Utilization of multiplicative LSTM to improve classification accuracy.

➢ *Different Approaches to HAR:*

A HAR system is necessary in order to accomplish the purpose of identifying human activity. For this reason, sensor-based and vision-based activity recognition are the two most widely utilized methods. We are able to categorize them.

- *Pose Based Approach:*

Pose-based approaches focus on analysing the human body's posture or pose to recognize activities. This approach often involves using computer vision techniques to extract key joint positions or body configurations from images or videos. One common technique in pose-based HAR is using pose estimation algorithms, such as OpenPose or PoseNet, to detect and track key points on the human body. These algorithms can estimate the positions of joints like elbows, shoulders, hips, and knees. Once the pose information is

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1530

obtained, machine learning or deep learning models can be trained to classify different activities based on the detected poses. Features derived from pose data, such as joint angles or body part velocities, can be used as input features for these models. Pose-based approaches are particularly useful when sensor data is not available or when activities can be reliably inferred from visual cues, such as in surveillance or human-computer interaction applications.

- *Sensor Based:*

Sensor-based approaches involve using various types of sensors, such as accelerometers, gyroscopes, magnetometers, or inertial measurement units (IMUs), to capture motion or physiological signals associated with human activities. These sensors can be embedded in wearable devices like smartwatches, fitness trackers, or smartphones, or they can be placed in the environment (e.g., on walls or floors) to capture activity data. Data collected from sensors typically include time-series measurements of acceleration, rotation, or other physical quantities. Machine learning algorithms, such as decision trees, support vector machines (SVMs), or deep neural networks, can then be trained on this data to recognize different activities. Sensor-based approaches are widely used in applications like activity tracking, fall detection, sports performance analysis, and healthcare monitoring due to their versatility and non-intrusiveness.

➢ *Proposed System*

Human Activity Recognition (HAR) is a burgeoning field with myriad applications, driven by the proliferation of devices equipped with sensors and cameras, and the rapid advancements in Artificial Intelligence (AI). HAR involves the art of identifying and naming various human activities using AI algorithms on data collected from these devices. The data from these sources, including smartphones, video cameras, RFID, and Wi-Fi, provides a rich source of information that can be harnessed for recognizing human activities.

The relationship between the growth of HAR and AI is symbiotic. As AI techniques evolve, they enable more sophisticated and accurate methods for extracting meaningful information from raw sensor data, giving rise to improved HAR systems. In the past, HAR models relied on single images or short image sequences. These advanced models, such as CNNs combined with LSTMs, have made HAR designs more robust and adaptable.

Human activities are diverse and can encompass a wide range of motions and positions. HAR has diverse applications, including healthcare, surveillance, monitoring, and assistance for the elderly and people with disabilities. It is imperative to note that while the field is making great strides, there are challenges that need to be addressed. These challenges include the selection of appropriate sensors, data collection, recognition performance, energy efficiency, processing capacity, and system flexibility.

Our goal in this project is to create an efficient and effective human recognition system that can process video and image data to detect tasks performed. The system is versatile and can find applications in many areas, such as caring for and helping people with special needs.

III. METHODOLOGY

This application harnesses MediaPipe's human pose estimation pipeline, powered by machine learning models like convolutional neural networks (CNNs) or similar architectures. These models enable real-time detection and estimation of key points on the human body, resulting in graphical representations of skeletal structures. These visuals serve as a roadmap of a person's posture and movements, offering valuable insights into their activities. At its core are pose landmarks, individual coordinates corresponding to specific joints in the body, pivotal for accurately depicting a person's pose and tracking their movements over time. Furthermore, the connections between these key points, known as pairs, contribute to a deeper understanding of human motion dynamics.

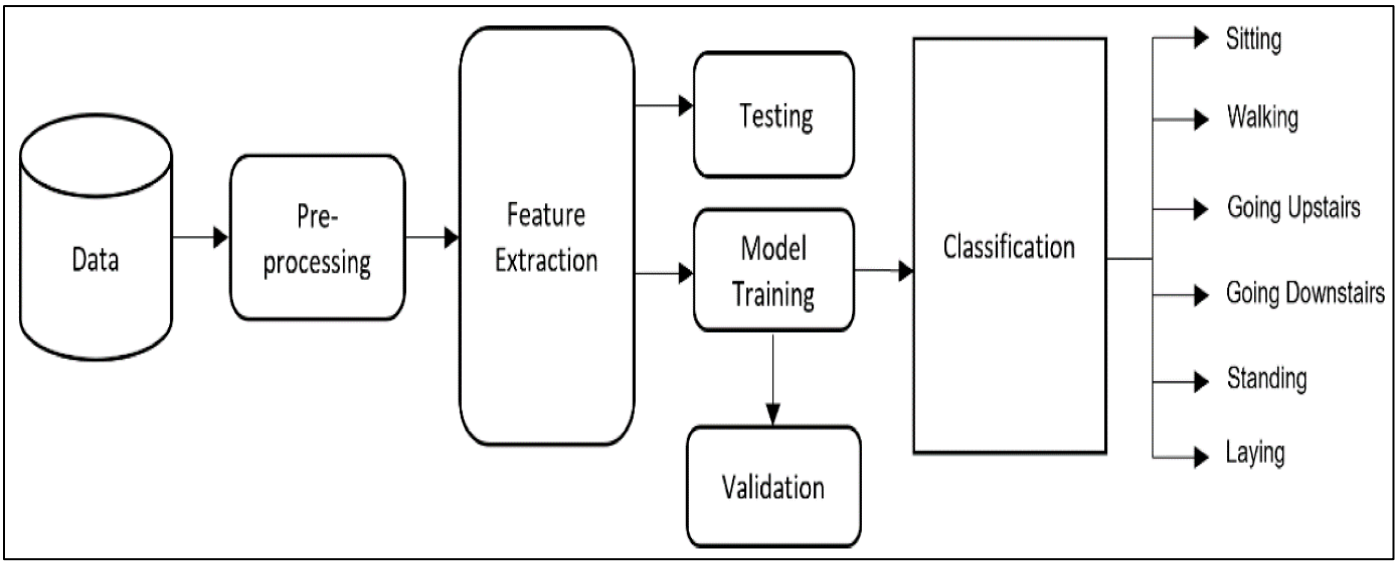

Fig 2 Conventional Machine Learning Architecture for Human Activity Recognition

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1530

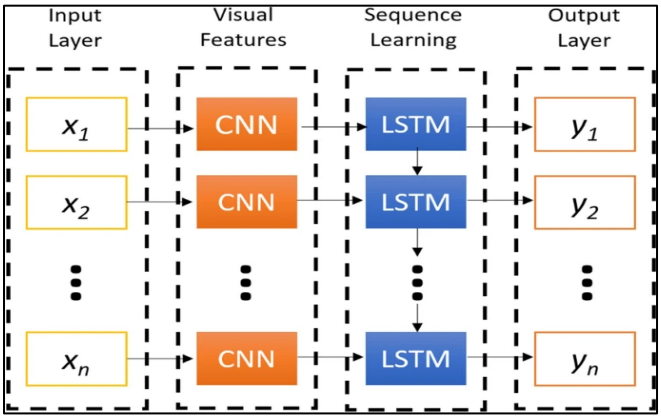

Fig 3 Basic Architecture of the CNN-LSTM Network

Another critical component of the HAR system is its user interface (UI), acting as the primary point of interaction between the user and the application. Leveraging the Tkinter library, we have crafted an intuitive and visually appealing UI that facilitates seamless navigation and interaction. The interface prominently features labeled buttons for various functionalities, neatly categorized under Camera and Video options, empowering users to effortlessly select their preferred mode for HAR.

➢ *Working*

The project boasts a Graphical User Interface (GUI) crafted with Tkinter, facilitating human tracking while leveraging Mediapipe's pose estimation model for exercise supervision. Through Tkinter, the GUI provides users with a variety of exercise tracking options, including fighting, smoking, walking, reading, and playing. It offers the flexibility of choosing between real-time camera recognition or video input, with distinct buttons to toggle between the two modes.

IV. RESULT

These are the results we achieved after developing the application. The result shows the GUI, Human Activity Recognition through real-time video feed and prerecorded video.

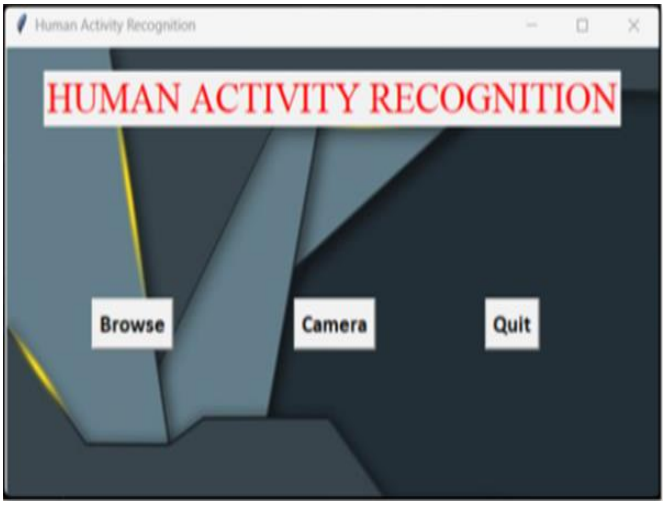

Fig 4 Graphical user Interface

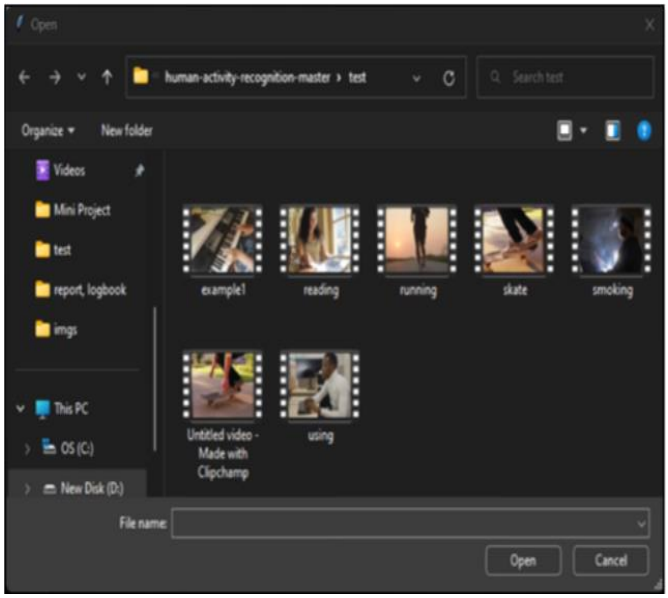

Fig 5 Selecting Video

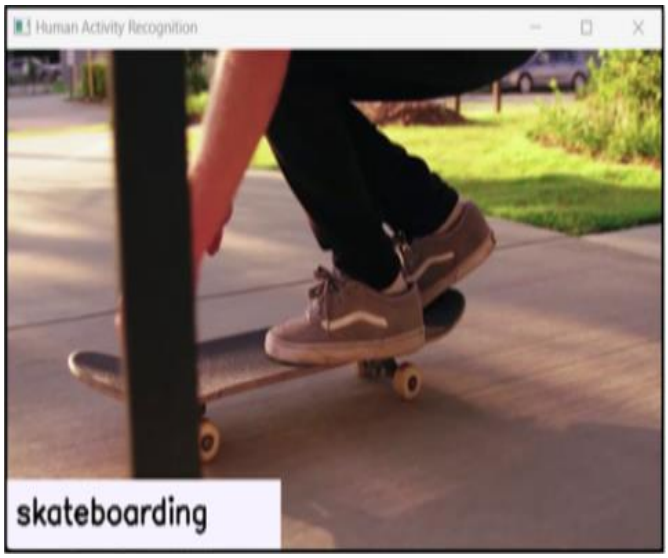

Fig 6 Video Output

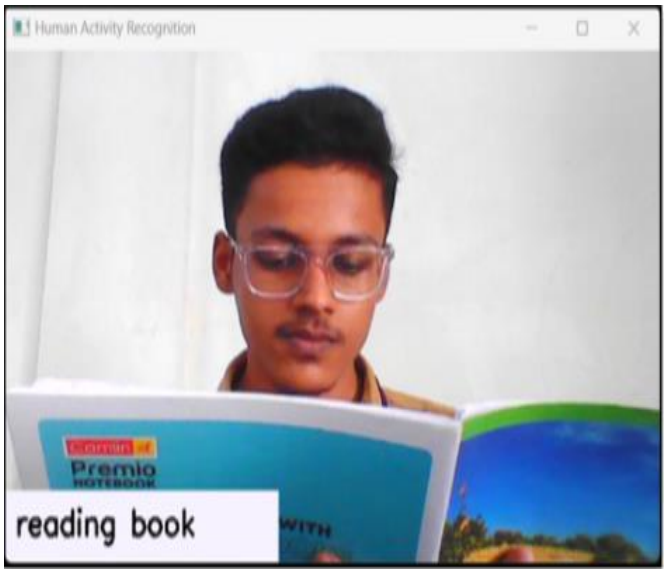

Fig 7 Real-Time Camera Output

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1530

V. FUTURE SCOPE

The future holds a wealth of opportunities for further enhancing and extending Human Activity Recognition systems. Here are some promising areas of development and exploration:

- *Multimodal Integration:*
  Incorporating data from diverse sensor types, such as audio, accelerometers, and more, can enable richer context and improved recognition accuracy.

- *Privacy and Ethics:*
  Researching and implementing robust privacy measures and ethical guidelines is essential for deploying HAR systems in sensitive or public environments.

- *Real-World Deployment:*
  Scaling up the usage of HAR systems in practical domains like smart homes, healthcare, and personalized assistance offers a wide array of benefits.

- *Edge Computing:*
  Investigating the feasibility of deploying HAR models on edge devices can lead to more responsive real-time applications without heavy reliance on cloud resources.

- *Incremental Learning:*
  Developing techniques for incremental learning allows systems to adapt to new activities and environments over time without complete retraining.

- *Environment Robustness:*
  Enhancing system robustness to challenging environments, including adverse weather and lighting conditions, is a crucial research direction.

- *User Interaction:*
  Exploring natural and intuitive user interaction methods, like gesture recognition or voice commands, can enhance the usability of HAR systems.

In conclusion, the future of Human Activity Recognition is filled with exciting possibilities. The continual development of these systems is set to unlock new levels of accuracy, versatility, and applicability in various domains. These advancements will contribute to the ongoing evolution of intelligent systems that can understand and respond to human actions across diverse scenarios, making a significant impact on society and industries.

VI. CONCLUSION

In an era dominated by the advancements in computer vision, the development and implementation of Human Activity Recognition (HAR) systems have emerged as a compelling and indispensable technology. These systems has proven a remarkably effective in addressing multitude of real-world applications, from surveillance and monitoring to providing invaluable assistance to elderly and visually impaired individuals. The project's CNN-LSTM-based HAR system represents a significant stride in this direction, offering a versatile and accurate approach to recognize and categorize a wide array of human activities. This system's adaptability to video streams and image data showcases its potential in addressing the diverse needs of various industries and domains.

REFERENCES

[1]. J. Zhang, A. Yang, C. Miao, X. Li, R. Zhang and D. N. H. Thanh, "3D Graph Convolutional Feature Selection and Dense Pre-Estimation for Skeleton Action Recognition," in IEEE Access, vol. 12, pp. 11733-11742, 2024, doi: 10.1109/ACCESS.2024.3353622.

[2]. M. Atikuzzaman, T. R. Rahman, E. Wazed, M. P. Hossain and M. Z. Islam, "Human Activity Recognition System from Different Poses with CNN," 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2020, pp. 1-5, doi: 10.1109/STI50764.2020.9350508.

[3]. M. Ramzan, A. Abid, M. Bilal, K. M. Aamir, S. A. Memon and T. -S. Chung, "Effectiveness of Pre-Trained CNN Networks for Detecting Abnormal Activities in Online Exams," in IEEE Access, vol. 12, pp. 21503-21519, 2024, doi: 10.1109/ACCESS.2024.3359689.

[4]. S. Win and T. L. L. Thein, "Real-Time Human Motion Detection, Tracking and Activity Recognition with Skeletal Model," 2020 IEEE Conference on Computer Applications(ICCA), Yangon, Myanmar, 2020, pp. 1-5, doi: 10.1109/ICCA49400.2020.9022822.

[5]. A. Gupta, K. Gupta, K. Gupta and K. Gupta, "A Survey on Human Activity Recognition and Classification," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020, pp. 0915-0919, doi: 10.1109/ICCSP48568.2020.9182416.

[6]. https://www.mdpi.com/2078-2489/13/6/275