



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: <https://doi.org/10.22214/ijraset.2024.65191>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Human Anomaly Detection System Using YoloV8 and LSTM

Jash Tandel¹, Siddhesh Darak², Khagesh Desai³, Meet Desai⁴, Mayank Kothari⁵

NMIMS University Department of Computer Engineering, Mumbai, India

Abstract: The “Human Anomaly Detection System” seeks to augment real-time surveillance by pinpointing atypical human activities via a hybrid deep learning strategy that merges RGB video frame analysis with pose estimation. By employing multi-stream neural networks such as YOLO for object detection and MobileLSTM for the classification of temporal actions this system utilizes attention mechanisms to uncover subtle anomalies in human behavior. The limitations of traditional surveillance methods, which usually rely on manual monitoring or rule-based frameworks that are prone to errors, are addressed by this research. Public safety, healthcare, and educational institutions can all benefit from the system’s notable improvements in recognizing complex, non-rigid human behaviors, despite the difficulties in combining spatial and temporal data. This technology is positioned as a powerful solution for anomaly identification in dynamic contexts since evaluation results show a high degree of accuracy and dependability. Future work will focus on improving model modularity and scalability to enable wider hardware flexibility and investigating richer datasets to improve performance across a range of scenarios.

Index Terms: Human Activity Anomaly Detection, Deep Learning, YOLO, MobileLSTM, Computer Vision, Real-Time Surveillance, Pose Estimation, Public Safety, Anomaly Detection System.

I. INTRODUCTION

Surveillance systems have become critical for efficient safety improvement in modern technological environments in both the public and private realms. Traditionally, anomaly detection techniques typically depend on rules-related methodologies or human intervention, and accuracy becomes a concern, particularly in sophisticated and real-time contexts. This calls for more evolved systems able to identify unusual behaviors quickly and accurately. Such a system aims to build a high-class anomaly detection system that will rely on HAR and HAAD techniques, thusly based on fundamental deep learning algorithms. In general, such a framework is expected to enhance security and safety in monitored environments, including public and corporate premises, as well as educational facilities, through the enhanced understanding of complex human behaviors.

Although this, the system is able to make autonomous identification of behavioral patterns through the usage of Computer Vision and techniques in deep learning, reducing the burden put on human operators and minimizing false positives. Despite all the possible challenges, the benefits this kind of system presents are considerable in improving surveillance processes and reinforcing strong comprehensive security protocols. However, their integration calls for careful examination from the ethical dimensions to the implementation challenges, since the success of those systems is sensitive to their design and implementation. The system architecture aims at extensibility along with computational efficiency, hence easily deployable on a variety of hardware configurations ranging from modest campus deployments to megapolis-wide surveillance systems. With real-time processing ability, anomalies get detected immediately, which further boosts response efficiency alongside productivity in operations. Therefore, this project proposes to provide an absolute leap in the field of anomaly detection. This will be achieved by developing a technically sound yet flexible framework to be applied towards various practical ends. This system has promise in crime prevention, emergency response, and general public safety improvement, coinciding with the swell of demand for automated and intelligent surveillance systems intended to promote societal welfare and security. However, a rigorous study of technical challenges underlies these goals.

II. LITERATURE REVIEW

Human Activity Anomaly Detection (HAAD) is of paramount importance as a domain of study in the field of computer vision, identification this time, between normal versus abnormal human activities. It is also critical for many applications such as surveillance, assisted living environment and human-machine interactions, robotics, and sports performance analysis. One major obstacle encountered in HAAD is the inconsistent definitions of anomalies that exist across various datasets and contextual settings. Because of these key aspects, this research is primarily focused on monitoring video feeds in surveillance scenarios and examines human activities denoted by semantic labels at the posture level.

This includes actions like running, bending, jumping and sitting. However, through the integration of Convolutional Neural Networks (CNNs) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for modeling temporal dynamics, this framework achieves a high degree of precision in identifying posture-related anomalies.

Comprehensive evaluations of widely utilized video datasets, in conjunction with newly developed custom datasets, illustrate the system's potential for recognizing both single-target and multi-target anomalies in practical environments, akin to CCTV systems. The proposed methodology integrates RGB and pose information for the simultaneous detection of human action anomalies, thus addressing prior limitations in pixel-level analysis. The SSD (Single Shot MultiBox Detector) is employed because it meets the requirements for real-time detection of objects and bounding boxes necessary to identify abnormal activities. Jagriti et al. [2] present a novel anomaly detection system within the analytical framework of close-circuit television footage aimed at enhancing safety measures. The study constructs the system based on a machine learning algorithm that learns patterns of normal behavior in the observed environment to detect anomalies as potential outliers, improving surveillance accuracy. However, this approach still faces challenges in diverse environmental conditions.

Ghazal et al. (3) explore whether video annotation related to human pose serves as a valid method for anomaly detection based on pose. There are several benefits associated with employing pose information (privacy concerns are mitigated) and the computational load is substantially lessened compared to conventional pixel-oriented methods. However, it also brings to light various challenges it encounters: for example, noisy skeleton data, excessive pixel information and a somewhat limited richness of features that may impede detection. Sedat et al. (4) introduce a novel approach that, when human pose is established, facilitates improved video analysis intended for the identification of pre-emptive crimes. This research particularly focuses on shoplifting and aims to anticipate events during critical incidents, as well as behaviors that could lead to criminal activities, such as looking, pacing and changes in direction.

Tran et al. (5) address the challenges involved in video anomaly detection, particularly in the realms of video surveillance and healthcare applications. The authors discuss the complexities that emerge from variability in anomalies, which is affected by differences among individuals, objects, or various contextual factors. Furthermore, they emphasize the problems arising from the limited availability of annotated datasets and the uneven distribution of training data. This paper introduces a novel approach for semantic-explanatory anomaly identification. It utilizes readily available object detection and individual identification techniques within video frames to extract semantic features, such as motion, appearance and posture, which are essential for recognizing anomalies. However, Caijian et al. (6) offer YOLO-ABD, a lightweight multi-scale detection model intended to improve public safety and surveillance by effectively identifying anomalous behavior among pedestrians. The model underwent evaluation with the IITB-Corridor test dataset, yielding notably impressive mAP scores: 89.3% at mAP50 and 60.6% at mAP50-95. However, the results suggest that while the model is effective, further improvements could be necessary to address specific limitations in real-world scenarios.

Gawande et al. (7) examine the difficulties associated with person detection and the recognition of suspicious activities within image-based surveillance systems. This analysis brings to light various concerns tied to individual movements and postures, alongside the complexities posed by the surrounding environment. They improved the YOLOv5 architecture by developing a method that effectively extracts movement features; these features elucidate aspects such as movement, speed and direction, ultimately aiming to enhance the detection of both global and local anomalous behaviors. Jain et al. (8) propose PoseCVAE, a novel methodology that employs generative modeling for anomaly detection specifically focused on human poses. This method integrates both self-supervised and unsupervised learning within a hybrid training framework, which fortifies the representations of the encoder and decoder pertaining to human pose trajectories. Huang et al. (9) introduce a significant innovation: the hierarchical graph-embedded pose regularity learning framework. This framework is designed to model intra-person joint relationships and inter-person interactions, thereby capturing more intricate contextual information present in surveillance scenes. However, the effectiveness of these approaches may vary, depending on the specific scenarios and environmental conditions encountered.

Bharadwaj et al. (10) propose a CNN-based system which enhances real-time anomaly detection. This underscores the significance of automation in contemporary surveillance; however, it also identifies key challenges that must be addressed. Although the system shows promise, further refinements are necessary because effective implementation is crucial for success.

Sanjeevkumar et al. (11) underscore the necessity of an automated Human Abnormal Activity Recognition (HAR) system intended to enhance societal safety. The proposed solution utilizes machine learning—specifically Convolutional Neural Networks (CNNs)—to process and analyze CCTV footage. Hobeom, Jeon et al. (12) present PASS-CCTV, an innovative surveillance system crafted to refine CCTV footage analysis, particularly in adverse environmental conditions. This system effectively confronts critical challenges in surveillance; it offers improved security through advanced algorithms, deep learning integration and user participation.

However, these developments raise questions regarding implementation and user trust, but they also pave the way for more effective monitoring techniques. Although challenges remain, the potential benefits of such systems are significant because they could lead to a safer environment.

III. OBJECTIVE

The creation of a dependable real-time anomaly detection system (as detailed in the sources) holds considerable potential for enhancing security, safety and efficiency across multiple sectors. This project aims to contribute significantly to society by utilizing the capabilities of deep learning and computer vision; however, it also seeks to advance the domain of Human Activity Anomaly Detection. Although challenges exist, the integration of these technologies is crucial because it can lead to more effective monitoring and response strategies. Thus, the implications of this work extend beyond mere technical achievements.

A. Main Objective

The primary goal of this project is to create a strong real-time anomaly detection system that can identify atypical human activities through a blend of deep learning and computer vision techniques. This system intends to improve detection accuracy across diverse monitoring situations by tackling the challenges presented by non-rigid human movements and addressing limitations in existing methods. It will utilize advanced deep learning algorithms to process and analyze intricate human behaviors, thus enabling accurate and real-time detection of anomalous activities in monitored environments (such as public spaces, corporate campuses and educational institutions). However, despite these advancements, there may still be challenges to overcome.

B. Significance

The importance of this project resides in its ability to meet several urgent requirements in the fields of security and surveillance. The system can be implemented across diverse environments (for example, public areas, transport centers and educational facilities) to identify potentially hazardous or suspicious behaviors. This capability can assist in crime prevention, ensuring swift reactions to emergencies and fostering safer surroundings. By automating the anomaly detection mechanism, the system diminishes the need for constant human oversight, which is frequently susceptible to mistakes and inefficiencies. This enables security personnel to concentrate on critical scenarios and enhances the overall efficacy of the surveillance system. Real-time anomaly detection facilitates immediate responses to possible threats or atypical occurrences, thereby minimizing the repercussions of such events and potentially preserving lives (see references 14, 15 and 16). Furthermore, this project advances the research and development of more resilient, efficient and flexible HAAD systems, thereby expanding the frontiers of computer vision and deep learning applications in practical situations.

IV. FEASIBILITY STUDY

A. Technical Feasibility

- 1) The project utilizes established and easily accessible technologies, such as deep learning frameworks (e.g., TensorFlow and PyTorch), computer vision libraries (e.g., OpenCV) and pose estimation tools (like OpenPose and MediaPipe).
- 2) The presence of robust hardware, including MacBook M1 and M2 chips, guarantees ample computational resources for real-time video data processing.
- 3) The sources furnish comprehensive descriptions of the system's architecture, software prerequisites and algorithms, thus demonstrating a clear grasp of the technical components involved.
- 4) The effective execution of various methodologies, including YOLOv8 and MoBiLSTM, further enhances the project's technical feasibility.
- 5) The incorporation of standard data manipulation and analysis tools, such as Matplotlib and Seaborn, aids in data visualization and analysis, which are crucial for model development and assessment.
- 6) Potential technical challenges arise, particularly concerning the limited modularity of certain libraries (like MediaPipe), which may complicate integration and diminish flexibility.

B. Operational Feasibility

- 1) The system's capability to identify anomalies in real-time, combined with its intuitive interface (which is user-friendly), renders it appropriate for a variety of environments, such as educational institutions, workplaces, hospitals and public spaces.
- 2) The decentralized architecture not only enhances reliability but also diminishes the likelihood of system failure.

- 3) The advantages of automated anomaly detection encompass improved surveillance efficiency, proactive responses to potential threats and enhanced public safety.
- 4) Because of the system's capacity to lessen dependence on human monitoring and deliver timely alerts, it proves operationally efficient.
- 5) The necessity for ongoing training and updates to datasets—essential for maintaining system accuracy and addressing evolving threats—along with the potential for false alarms, must be taken into account and mitigated through meticulous system design and configuration.
- 6) Although these operational considerations exist, the extensive range of potential applications, the advantages of real-time detection and the emphasis on user-friendliness indicate the operational viability of the Human Anomaly Detection System.

V. SYSTEM ARCHITECTURE

In this research paper, two different approaches were explored and compared for the system architecture:

A. Approach-1 (Using YOLOv8 Model)

- 1) **Data Collection:** Roboflow was used to annotate the various human actions of Violence, Knife-wielding, Falling, Mounting and Walking in the preparation of data for the YOLOv8 model. The video was segmented into frames, which were step-by-step annotated with bounding boxes for each action. The output was saved to a text file with the exact spatial coordinates and their corresponding activity labels, as shown in Fig. 1. This systematic approach ensures a rightly labeled training dataset and therefore the soundly established development of action detection model.



Fig. 1. Annotated frames for human actions in YOLOv8 model

- 2) **Architectural Workflow:** The system architecture in Figure 2 to be covered in this paper explains a holistic approach to human action recognition and alerting using YOLOv8 and an SVM classifier. The whole process simply begins from an input video dataset, which undergoes preprocessing for improved data quality. Then, feature selection is carried out in order to identify attributes that are most relevant for classification purposes. The YOLOv8-based feature YOLO version 8 object detection for specific activities removes the human figure from background contents. The SVM classifier then categorizes the detected actions as violence or normal behavior. The alerting system is finally used to trigger notification for specific actions that could enhance real-time response capabilities.

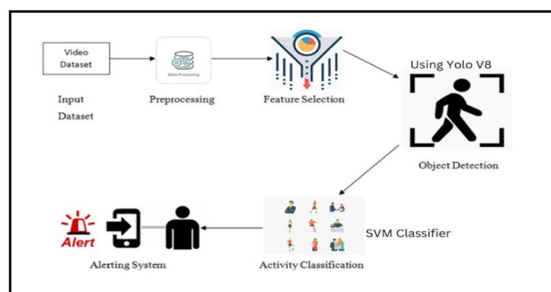


Fig. 2. Architectural Block Diagram- For YOLOv8 Classifier

- 3) *Advantages:* One of the most marked advantages of using YOLOv8 for Human Activity Recognition is the speed and real-time capabilities combined with high accuracy object detection. The architecture-based detection by YOLOv8 allows it to identify multiple human actions within a video frame, a capacity that suits security monitoring or detection when the response must be immediate. In addition, increasing the object detection accuracy of YOLOv8 ensures reliable activity recognition, especially if used with other classifiers such as SVM for activity classification into pre-defined actions.
- 4) *Disadvantages:* The major limitation of YOLOv8 for the task of HAR is that it cannot handle sequential data. Since YOLOv8 is mainly designed for object detection, it lacks the possibility to capture temporal patterns in video frames, which is a very important property in sequential action classification. This can lead to relatively low accuracy in movement recognition, such as the correct distinction between walking, running, or falling. To address this issue, YOLOv8 often has to be used in combination with another algorithm, probably in the form of RNNs or LSTM networks, which increases complications in the system and raises further requirements related to computational power.

B. Approach-2 (Using MediaPipe and MobileLSTM)

- 1) *Data Collection:* Mediapipe was utilized to extract skeletal keypoints from various human actions (such as Jump, Punch, Run, Sit, Squat, Stand, Walk, Wave and Kick) for training the MobileLSTM model. Initially, video data were converted into sequential frames; from these frames, skeletal coordinates were extracted on a frame-by-frame basis. The skeletal keypoint data for each frame was stored in CSV format, along with their corresponding activity labels. Approximately 12,000 images were processed and annotated, covering nine distinct action classes. This methodical approach guarantees (that) a properly structured sequential dataset is available for the effective development of the human action recognition model. However, challenges may occur during the annotation process, because meticulous attention to detail is essential. Although the process is demanding, it ultimately contributes to more accurate model performance.

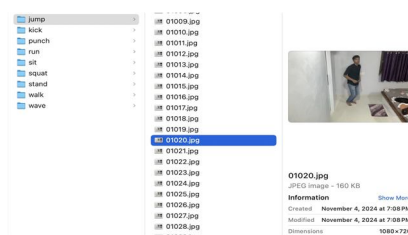


Fig. 3. Dataset Prepared for Mediapipe Feature Extraction

- 2) *Architectural Workflow:* The Architectural diagram illustrates a very comprehensive data processing pipeline. At the top, the "Configuration" section describes the system configuration as a whole, including all the "System Configuration" parameters that specify which data is moved and processed where the core libraries section describes the basic modules forming the backbone of the system. I/O Handler and Data Formatter modules are responsible for data input and formatting; the Skeleton Detector that separates skeletal structures from the data available; People Tracker responsible for identification and tracking of persons; Feature Extractor, which determines features of interest in the data; and Activity Classifier determines the activity or behavior. The main core component of the system will be "Main Processing Pipeline," which carries out fundamental data processing tasks. It starts at "Skeleton Extraction," where the system outlines skeletal forms from the input data and moves through to "Data Consolidation," where the data extracted is systematically organized and prepared for further analysis. This step is termed as "Feature Processing," wherein feature extraction techniques are applied to find out the most relevant features or patterns of interest in the aggregated data. These identified features form the basis for the "Model Training" step, which trains a machine learning or deep learning model upon to identify and classify the patterns in the data. Finally, the "Testing & Evaluation" part tests how effective the learned model is and provides significant insights regarding its precision and dependability concerning the task defined. At the end of the chapter, brief information is provided regarding the input and output data flows. Here, the input data are identified for the training images and for the validation information, which the system processes to yield the final product: the trained model along with its predictions. Here, in detail, is described system architecture that forms a robust data processing pipeline that draws heavily from many fundamental libraries and components to extract, integrate, and process data for assessment. This process results in a trained model that makes predictions or classifications based upon the input data. The detailed explanation given here may benefit as an important accompaniment to your research paper, wherein all the principal components are explained adequately and their role in the entire system is aptly described.

- 3) *Advantages:* One of the most notable advantages of employing MobileLSTM for Human Activity Recognition is its proficiency in effectively capturing temporal dependencies and sequential patterns inherent in human movements. The LSTM-based architecture allows the model to sustain temporal context across video frames. This characteristic makes it particularly adept at differentiating between similar actions, which may vary in their temporal execution (e.g., running versus walking). Furthermore, the lightweight nature of MobileNet, serving as the backbone network, ensures computational efficiency while maintaining a reasonable level of accuracy; thus, it is suitable for deployment on devices with limited computational resources. Moreover, the model's intrinsic ability to handle sequential data allows it to identify complex activities that unfold over time (1). This functionality is especially useful as it eliminates the need for additional elements of temporal processing. However, the system still has a potential for malfunction in certain contexts- this is important because it raises possibilities of possible limitations that may be present. Though the model achieves competence it is important to move forward with caution regarding the context in which it performs.
- 4) *Disadvantages:* Perhaps the most critical limitation of this MobileLSTM approach is that it needs identical skeletal extraction profiles across different frames. This is important because the network is dependent on the availability of the skeletal keypoints; therefore, any kind of errors or failures in the pose estimation could have a domino effect and culminate in being reflected in the final classification outcomes. Nevertheless, this methodology might face difficulties in scenarios that include several individuals or obstructions, given that these elements can hinder precise skeleton extraction. While the model demonstrates competence in identifying temporal patterns, it may not perform optimally when it comes to discerning intricate spatial details or nuanced postural changes. Such subtleties could be essential for distinguishing particular actions. Furthermore, it is very likely that the sequential processing characteristic has a small latency in recognition compared with fully spatial methods, which would plausibly influence real-time applications.

VI. APPLICATION

A. Public Safety

Real-time monitoring is crucial for quickly identifying and alerting authorities to potentially dangerous behaviors, such as fights, unattended items, or unusual crowd movements, particularly in crowded areas like train stations and airports. Improved security measures at public events (such as sporting events and concerts) make it easier to identify any suspicious or unusual activity quickly, which is crucial because it helps to avert incidents. Nevertheless, there are obstacles in the way of successfully implementing these steps. The early identification of aggressive or illegal behavior improves crime prevention and helps law enforcement keep the peace. These systems must be used with caution, though, as they may give rise to privacy issues. Despite its effectiveness, striking a balance between personal freedoms and security is still crucial in these situations.

B. Educational Institutions

The detection of bullying, fights, or unauthorized activities on school and university campuses is crucial; it promotes a safer environment (for both students and staff). Monitoring for early signs of emergencies, such as fires or other hazards, is essential because it improves response time and enhances student safety during critical situations. Furthermore, the observation of behavior in high-risk lab environments or workshops is necessary. This allows for the identification of actions that deviate from established safety protocols, which helps prevent accidents. However, it is important to remain vigilant, as even minor oversights can lead to significant consequences.

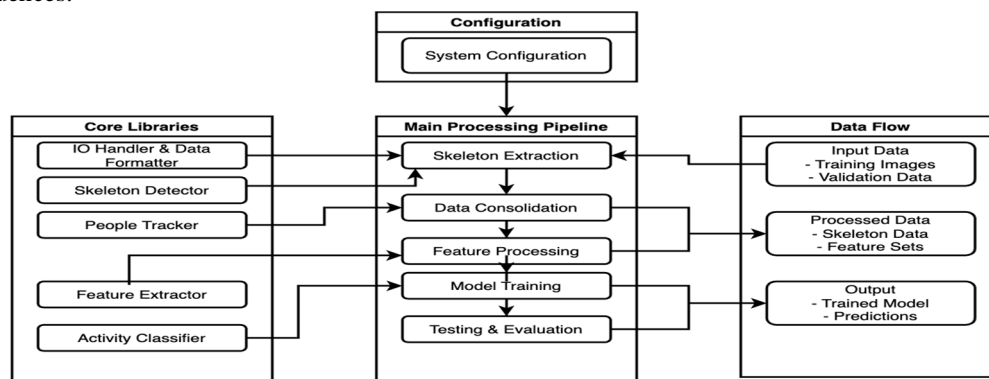


Fig. 4. Architectural Block Diagram - For MobileLSTM Based Classifier

C. Corporate and Industrial Security:

Improved workplace safety in (factories or warehouses) involves identifying hazardous behaviors or unauthorized access in restricted areas. Monitoring employee activity is crucial for compliance with safety standards; this is particularly important in high-risk environments such as chemical plants or construction sites. Detection of anomalies can indicate potential theft, fraud, or sabotage; however, it serves to protect assets and intellectual property. Although these measures are essential, they also require constant vigilance and adaptation, because the nature of risks continually evolves.

VII. RESULT

In this investigation, two methodologies were utilized for action recognition: YOLOv8 and MoBiLSTM. Each model underwent training and evaluation using a dataset that consists of six distinct action categories: walk, punch, push, run, kick and jump, with annotations developed in Roboflow. The YOLOv8 model was optimized for spatial classification, attaining an overall accuracy of 90%. Specific actions such as violence, mount and walking were classified with an accuracy of 80%, thus underscoring YOLOv8's effectiveness in single-frame spatial detection. However, its principal limitation resides in its inability to capture temporal dependencies, which renders it less suitable for real-time sequential activity classification. The classification outcomes for YOLOv8 (Approach 1) are illustrated in Table 1; this provides a comprehensive overview of its performance metrics.

In contrast, the MoBiLSTM model was developed to address the need for temporal recognition; it was trained on a 23-minute video that was pre-processed into frames and categorized into the same six classes. MoBiLSTM achieved a higher overall accuracy of 97%. Furthermore, the classification report (as shown in Table 2), along with confusion matrix results in Figure 6, indicates high accuracy across all action classes. The model's ability to capture frame-by-frame temporal relationships makes it particularly adept at identifying sequences—such as running and punching—where action continuity is critical. However, despite these strengths, challenges were encountered because of limited modularity in libraries like MediaPipe, which complicated integration. Additionally, the high computational complexity of MoBiLSTM may impact its feasibility for real-time use on devices with limited processing capabilities.

Precision is the ratio of correctly predicted data points to total predicted data points, defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

where TP and FP denote true positive and false positive, respectively.

Recall (Sensitivity) measures the proportion of correctly classified data points belonging to a particular class out of all data points in that class, calculated as:

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \quad (2)$$

F1-score measures the overall performance of a model by taking the harmonic mean of precision and recall, calculated as:

$$\text{F1-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

TABLE I
TABLE 1: CLASSIFICATION REPORT (APPROACH-1)

Class	Images	Instance	Precision	Recall	F1-Score
All	1800	140	0.875	0.914	0.911
Violence	400	400	0.978	1.000	0.995
Fall	190	190	1.000	0.997	0.995
Knife	310	310	0.639	0.570	0.576
Mount	380	380	0.977	1.000	0.995
Walking	120	120	0.783	1.000	0.995

TABLE II
TABLE: CLASSIFICATION REPORT (APPROACH-2)

Class	Images	Instances	Precision	Recall	F1-Score
All	11202	11000	0.96	0.96	0.97
Kick	1162	1150	0.97	0.96	0.96
Squat	964	930	0.99	0.99	0.99
Stand	1703	1500	0.99	0.96	0.96
Punch	799	799	0.99	1.00	0.99
Sit	1908	1600	0.99	0.99	0.99
Walking	1220	1200	0.99	0.96	0.92
Jump	1174	1170	0.99	0.99	0.99
Run	1033	1033	0.97	0.96	0.96
Wave	1239	1200	0.97	0.99	0.98

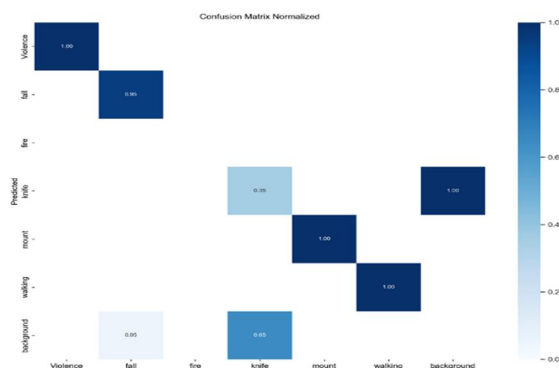


Fig. 5. Confusion Matrix (Approach 1)

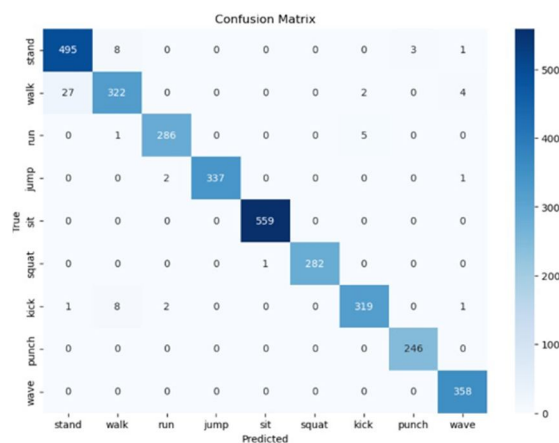


Fig. 6. Confusion Matrix (Approach 2)

VIII. CONCLUSION

This project explored two methodologies for detecting human anomalies: YOLOv8 and MoBiLSTM. The objective was to develop a system that can accurately identify abnormal human behavior in real-time. YOLOv8, although effective in object detection tasks, encountered limitations when utilized for sequential action classification (because it could not adequately capture temporal relationships between video frames). Conversely, the MoBiLSTM model excelled in this domain, effectively capturing temporal dependencies and demonstrating superior performance in both real-time execution and evaluation across test datasets. Consequently, MoBiLSTM emerged as a more appropriate model for human anomaly detection, particularly in recognizing intricate actions that involve multiple sequential movements. However, the challenges faced by YOLOv8 highlight the importance of selecting the right approach based on the specific requirements of the task at hand.

However, MoBiLSTM reveals areas requiring further enhancement. The integration of specific libraries (such as MediaPipe) proved challenging due to limited modularity, which consequently restricted flexibility within the model. Additionally, optimization is necessary to enhance MoBiLSTM's real-time performance across a broader array of hardware setups.

Addressing these challenges will improve the model's robustness, adaptability and scalability; this ultimately contributes to a more versatile solution for human anomaly detection, with applications spanning various fields. Although the current framework shows promise, it confronts several obstacles that must be surmounted.

IX. FUTURE SCOPE

The future scope of this project emphasizes the enhancement of modularity, the upgrading of libraries and the incorporation of advanced models for improved real-time anomaly detection. To create a more adaptable and scalable solution, future iterations of the project should prioritize modular development. This strategy will enable the incorporation of upgrade-able libraries, guaranteeing the system's flexibility, and it can accommodate new developments without requiring significant reconfiguration. It is simpler to update the model's components and maintain alignment with the most recent advancements in human anomaly detection technologies when the system is designed in this way. However, the intricacy of preserving such flexibility may provide difficulties. Despite the significant advantages, the possible drawbacks must be carefully considered. An important addition to the model architecture is the incorporation of YOLOv11 (a future version), which is expected to provide significant improvements in detection efficiency and speed.

Real-time applications greatly benefit from YOLOv11's remarkable anomaly detection speed of 10 milliseconds per frame, which reduces latency and improves responsiveness. For real-world applications like security monitoring, health care, and autonomous surveillance, where quick identification and reaction to unusual activity are critical, this increased speed is essential. To ensure that the system functions effectively in both high-power and resource-constrained contexts, more optimization and testing across several hardware platforms will be necessary. Making use of YOLOv11's speed and its modular architecture, which enables simple library upgrades, makes it possible to create a reliable, high-performing solution for human anomaly detection with little need for redesign.

REFERENCES

- [1] Yang, Yuxing, Federico Angelini, and Syed Mohsen Naqvi. "Pose-driven human activity anomaly detection in a CCTV-like environment." *IET Image Processing* 17.3 (2023): 674-686.
- [2] A, S, Kasthuri., Arvind, Balamurali., Arvind, Srinivasan., P., S., Lal, Priya. (2024). 1. Jagriti via Mudra: A pose based surveillance anomaly detection system. doi: 10.1109/accai61061.2024.10602098
- [3] Ghazal, Alinezhad, Noghre., Armin, Danesh, Pazho., Vinit, Katariya., Hamed, Tabkhi. (2023). 3. Understanding the Challenges and Opportunities of Pose-based Anomaly Detection. doi: 10.1145/3615834.3615844. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [4] Sedat, Kilic., Mihran, Tuceryan. (2023). Crime Detection from Pre-crime Video Analysis with Augmented Pose Information. 1-6. doi: 10.1109/eit57321.2023.10187325
- [5] Minh-Hanh, Tran. (2023). Semantic Features Extraction for Anomaly Detection from Real-World Videos*. 183-188. doi: 10.1109/rivf60135.2023.10471872
- [6] Caijian, Hua., Kun, Luo., WU, Ya., Rui, Shi. (2024). 4. YOLO-ABD: A Multi-Scale Detection Model for Pedestrian Anomaly Behavior Detection. doi: 10.20944/preprints202407.0522.v1
- [7] [7] Ujjwala, Gawande., Kamal, Hajari., Yogesh, Golhar. (2024). 5. Novel person detection and suspicious activity recognition using enhanced YOLOv5 and motion feature map. *Artificial Intelligence Review*, doi: 10.1007/s10462-023-10630-0.
- [8] Yashswi, Jain., Ashvini, Kumar, Sharma., Rajbabu, Velmurugan., Biplab, Banerjee. (2021). PoseCVAE: Anomalous Human Activity Detection. doi: 10.1109/ICPR48806.2021.9412132.
- [9] Chao, Huang., Yabo, Liu., Zheng, Zhang., Chengliang, Liu., Jie, Wen., Yong, Xu., Yaowei, Wang. (2022). Hierarchical Graph Embedded Pose Regularity Learning via Spatio-Temporal Transformer for Abnormal Behavior Detection. doi: 10.1145/3503161.3548369.
- [10] M., V., Bharadwaj., Nekkanti, Harshini., Prathapani, Mounika., M, Sowmya. (2022). Human abnormal activity recognition. *International journal for innovative engineering and management research*, 491-499. doi: 10.48047/ijiemr/v11/i06/30.
- [11] Sanjeevkumar, Angadi., Chellapilla, V., K., N., S., N., Moorthy., Mukesh, Kumar, Tripathi., Bhagyashree, Ashok, Tingare., Sandeep, Kadam., Kapil, Misal. (2024). An NFMF-DBILSTM model for human anomaly detection system in surveillance videos. *Indonesian Journal of Electrical Engineering and Computer Science*, 36(1):647-647. doi: 10.11591/ijeecs.v36.i1.pp647-656.
- [12] Hobeom, Jeon., Hyung-Min, Kim., Dohyung, Kim., Kim, Jj. (2024). PASS-CCTV: Proactive Anomaly Surveillance System for CCTV footage analysis in adverse environmental conditions. *Expert Systems with Applications*, doi: 10.1016/j.eswa.2024.124391.
- [13] Yuting, Qiu., James, Meng., Baihua, Li. (2024). Automated Falls Detection Using Visual Anomaly Detection and Pose-based Approaches: Experimental Review and Evaluation.
- [14] Angelini, Federico. Novel methods for posture-based human action recognition and activity anomaly detection. Diss. Newcastle University, 2020.
- [15] Zheng, Ce, et al. "Deep learning-based human pose estimation: A survey." *ACM Computing Surveys* 56.1 (2023): 1-37.
- [16] M., Pallewar., Vijaya, R, Pawar., Arun, N., Gaikwad. (2024). Unusual Human Behavior Analysis Using the Deep Learning. doi: 10.1109/esci59607.2024.10497371.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)