VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Project Report on

Human Activity recognition for Pedestrian Safety

In partial fulfillment of the Fourth Year (Semester–VIII), Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2023-2024

Submitted by
Sahil Talreja (D17C-66)
Varun Salvi(D17C-55)
Siya Doshi(D17C-18)
Roshni Jaisinghani(D17A-24)

Project Mentor
Mrs. Vidya Zope

Professor, Computer Engineering
(2023-24)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Certificate

This is to certify that **Sahil Talreja,Varun Salvi,Roshni Jaisinghani,Siya Doshi** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "**Human Activity Recognition for Pedestrian Safety**" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor **Mrs.Vidya Zope in the year 2023-24 .**

This thesis/dissertation/project report entitled **Human Activity Recognition for Pedestrian Safety** by Sahil Talreja,Varun Salvi,Roshni Jaisinghani,Siya Doshi is approved for the degree of *B.E.*

Programme Outcomes	**Grade**
PO1,PO2,PO3,PO4,PO5,PO6,P O7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide: Mrs.Vidya Zope

Project Report Approval For (Computer Engineering)

This thesis/dissertation/project report entitled **Human Activity recognition for Pedestrian Safety** by **Sahil Talreja,Roshni Jaisinghani,Siya Doshi,Varun Salvi** is approved for the degree of **B.E Computer Engineering**.

Internal Examiner

External Examiner

Head of the Department

Principal

Date:
Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)
Sahil Talreja D17C - 66

(Signature)
Varun Salvi D17C - 55

(Signature)
Roshni Jaisinghani D17A-24

(Signature)
Siya Doshi D17C - 18

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mrs. Vidya Zope** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair** , for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E. PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

Abstract

Pedestrian safety is a pressing concern in today's urban landscapes,exacerbated by factors such as distracted driving and non-compliance with traffic regulations. This project proposes the implementation of Human Activity Recognition (HAR) techniques to address these challenges effectively. HAR involves the detection and classification of human activities using video analysis and Machine Learning algorithms, offering real-time insights into pedestrian behaviors.

The primary goals of this initiative include prioritizing the safety of vulnerable groups such as the elderly and children and developing a user-friendly interface for various stakeholders. By accurately identifying and categorizing human activities, the HAR system enables proactive safety measures and timely interventions to mitigate potential risks to pedestrians.

Urban areas witness a surge in pedestrian accidents, emphasizing the urgency of deploying intelligent systems to interpret pedestrian intentions and behaviors near roadways. Through HAR technology, we aim to empower stakeholders to collaborate effectively in ensuring pedestrian safety and creating safer urban environments for all.

The implementation of HAR techniques for pedestrian safety holds immense potential in mitigating the risks associated with urban pedestrian environments. By prioritizing the safety of vulnerable groups and fostering collaboration among stakeholders, we can work towards creating safer and more pedestrian-friendly urban landscapes.

Chapter 1: Introduction

1.1 Introduction to the project

In our rapidly evolving urban landscapes, the safety of pedestrians stands as a paramount concern. As our cities grow denser and more bustling, the need for innovative solutions to safeguard individuals traversing these complex environments becomes increasingly pressing. One such solution lies in the realm of cutting-edge technologies, notably Human Activity Recognition (HAR), which holds the potential to revolutionize pedestrian safety.

HAR encompasses the real-time identification and categorization of human movements and behaviors through the utilization of video clips and sophisticated algorithms. This project embarks on an exploration of the multifaceted implications of HAR for pedestrian safety, delving into the intricate workings of the algorithms and machine learning techniques that underpin its functionality. By meticulously analyzing and understanding these mechanisms, the project endeavors to ensure that HAR systems can reliably distinguish between routine activities and potential safety hazards, thereby enhancing their efficacy in safeguarding pedestrians.

Moreover, the discourse extends beyond theoretical considerations to delve into the practical applications of real-time HAR in various domains, including urban planning, driver assistance systems, emergency response protocols, and beyond. By illuminating the diverse array of contexts in which HAR can be leveraged to bolster pedestrian safety, the project aims to foster a deeper appreciation for its transformative potential.

Furthermore, the project underscores the importance of seamless integration, emphasizing the need to incorporate HAR technologies into existing infrastructure and systems. By forging synergistic connections between technology and safety measures, a new paradigm of urban resilience can be cultivated, wherein innovative technologies serve as invaluable allies in the ongoing quest to prioritize pedestrian well-being amidst the dynamic tapestry of urban life.

1.2 Motivation

The drive to enhance pedestrian safety through Human Activity Recognition (HAR) is fueled by a compelling urgency and a series of potent motivations. Pedestrian accidents are far too common in urban areas globally, resulting in injuries, fatalities, and significant social and economic costs. The urgent need to mitigate these tragedies inspires us to explore innovative solutions like HAR, which has the potential to significantly reduce the frequency and severity of pedestrian accidents by enabling real-time monitoring and

response to pedestrian behaviors. Additionally, as cities continue to grow and evolve, the challenges associated with pedestrian safety become increasingly complex. HAR offers a dynamic and adaptable approach to address these challenges, ensuring that safety measures keep pace with the rapid urbanization and changing dynamics of cityscapes. Moreover, enhancing pedestrian safety not only prevents accidents but also enhances the overall quality of life in urban environments. By creating safer streets and walkways, HAR contributes to fostering vibrant, inclusive, and livable cities where residents can navigate with confidence and peace of mind. Furthermore, pedestrian safety is a universal concern, transcending geographical boundaries. By leveraging HAR technology, we not only strive to improve pedestrian safety locally but also contribute to the global community by sharing knowledge, best practices, and innovative solutions to create safer urban spaces worldwide.

1.3 Problem Definition

The aim of the project is to leverage Human Activity Recognition (HAR) technology to dramatically enhance pedestrian safety within urban environments. By implementing HAR techniques, the project seeks to proactively identify and mitigate risks associated with pedestrian accidents. Through real-time monitoring and analysis of pedestrian behaviors, the project aims to create a safer urban environment where pedestrians can navigate confidently and securely.

The primary objective is to dramatically enhance pedestrian safety within urban environments. This pressing issue arises from the rapid urbanization of cities, resulting in higher population density and increased foot traffic. As a consequence, pedestrian accidents are on the rise, with outdated infrastructure, distracted behaviors, and other factors exacerbating the problem. Leveraging emerging technologies such as Human Activity Recognition (HAR) presents an opportunity to proactively identify and mitigate risks. With a legal and ethical imperative to protect citizens and a desire to improve the overall quality of life in cities, this multi-faceted approach aims to create an urban environment where pedestrians can move confidently and securely.

1.4 Existing Systems

1.4.1 A Large-Scale Benchmark for 3D Human Activity Understanding

In this system, it presents a vast dataset for RGB+D human action recognition, comprising over 114,000 video samples and 8 million frames from 106 subjects. It encompasses 120 diverse action classes. It assess existing 3D activity analysis methods

and demonstrate the advantages of deep learning in this context. It also address one-shot 3D activity recognition, proposing an effective Action-Part Semantic Relevance-aware (APSR) framework. This dataset facilitates the development of data-intensive learning techniques for human activity understanding

1.4.2 Action recognition based on 2D skeletons extracted from RGB videos
This presents a novel approach for action recognition using RGB videos, capitalizing on deep learning advancements. It transforms skeletal motion data into 2D images, enabling action recognition solely from RGB videos. A deep neural network, OpenPose, extracts 18-joint skeletons from RGB video frames, which are then encoded into RGB channels. Various encoding methods were explored, and deep neural networks designed for image classification, such as ResNet, achieved impressive accuracy on the NTU RGB+D database, outperforming many state-of-the-art results (83.317% cross-subject and 88.780% cross-view accuracy)

1.4.3 Human Activity Recognition Using Acceleration Data from Smartphones
This is an android and IOS app that introduces a new dataset for human activity recognition and fall detection using smartphone-acquired acceleration samples. The dataset, comprising 11,771 samples from 30 subjects, categorizes activities into 17 fine-grained classes grouped into two coarse-grained classes: activities of daily living (ADL) and falls. It is meticulously annotated, enabling sample selection based on criteria such as activity type, age, and gender. The dataset is rigorously benchmarked with different classifiers and feature vectors, revealing that distinguishing falls from ADLs is relatively straightforward, but identifying specific fall types presents challenges due to similar acceleration patterns. The inclusion of the same subject in training and test data enhances classifier performance, emphasizing individual variability in activity execution.

1.5 Lacuna of the existing systems

Real-World Complexity : Many existing Human Activity Recognition (HAR) systems struggle to accurately recognize human activities in complex real-world scenarios, such as crowded urban environments, adverse weather conditions, or mixed traffic situations.

Cross-Domain Adaptation : HAR systems are often designed for specific domains (e.g., surveillance, sports), and adapting them for pedestrian safety in diverse settings can be challenging.

Privacy Concerns : Balancing the need for safety with privacy concerns is essential. Some systems may involve the collection of sensitive data, raising questions about data privacy and consent.

Anomaly Detection : Identifying unusual or unexpected human activities (anomalies) is crucial for safety but often neglected in traditional HAR systems.

Scalability : Scaling HAR systems to cover larger areas or cities and handling a massive influx of data can be a logistical and computational challenge.

1.6 Relevance of the Project

The project is profoundly relevant due to the escalating importance of pedestrian safety in urban environments. As cities expand and pedestrian traffic grows, the project addresses a pressing concern by harnessing technologies like Human Activity Recognition (HAR). It not only directly impacts lives by reducing accidents but also aligns with contemporary smart city initiatives, emphasizing safety and efficient urban planning. Moreover, it has economic implications, lessening healthcare costs and boosting productivity. Promoting pedestrian safety supports environmental sustainability, encouraging eco-friendly transportation. Ethically, it fulfills the duty of governments to protect citizens, ultimately enhancing the quality of life for urban residents and making our cities safer and more liveable.

Chapter 2: Literature Survey

2.1 Research Papers Referred

1)Jun Lui ,Gang Wang(2019):NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding .IEEE Transactions on Pattern Analysis and Machine Learning

Abstract: In this research, we present a vast dataset for RGB+D human action recognition, comprising over 114,000 video samples and 8 million frames from 106 subjects. It encompasses 120 diverse action classes. We assess existing 3D activity analysis methods and demonstrate the advantages of deep learning in this context. We also address one-shot 3D activity recognition, proposing an effective Action-Part Semantic Relevance-aware (APSR) framework. This dataset facilitates the development of data-intensive learning techniques for human activity understanding.

Inference:This study highlights the limitations of current 3D action recognition benchmarks, such as small subject diversity, limited action categories, fixed camera viewpoints, constrained environments, and a shortage of video samples. To address these shortcomings, the researchers introduce a large-scale benchmark dataset called NTU RGB+D 120. It contains 114,480 RGB+D video samples from 106 human subjects, diverse camera viewpoints, and a wide age range. This dataset aims to foster the development of more robust and data-driven 3D human activity analysis methods.

2) **Sophie Aubry, Sohaib Laraba*, Joëlle Tilmanne(2019).Action recognition based on 2D skeletons extracted from RGB videos

Abstract:This paper presents a novel approach for action recognition using RGB videos, capitalizing on deep learning advancements. It transforms skeletal motion data into 2D images, enabling action recognition solely from RGB videos. A deep neural network, OpenPose, extracts 18-joint skeletons from RGB video frames, which are then encoded into RGB channels. Various encoding methods were explored, and deep neural networks designed for image classification, such as ResNet, achieved impressive accuracy on the NTU RGB+D database, outperforming many state-of-the-art results (83.317% cross-subject and 88.780% cross-view accuracy).

Inference:The importance and challenges of action recognition are highlighted. The diverse applications of action recognition span multiple fields, from computer science to healthcare, demonstrating its wide-ranging significance. The challenges arise from the variability in movements, complexity of motion capture, and the need for realistic databases. Actions can be performed in numerous ways, influenced by context, individuals, and other factors, making it difficult to define action features. Capturing and

representing these movements require different systems, and amassing a substantial amount of data is essential, particularly for training neural networks. However, the creation and labeling of representative databases are time-consuming tasks. To achieve effective action recognition, it's crucial to address these challenges and develop comprehensive, realistic databases.

3)Daniela Micucci,Marco Mobilio(2017).A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones.

Abstract:This article introduces a new dataset for human activity recognition and fall detection using smartphone-acquired acceleration samples. The dataset, comprising 11,771 samples from 30 subjects, categorizes activities into 17 fine-grained classes grouped into two coarse-grained classes: activities of daily living (ADL) and falls. It is meticulously annotated, enabling sample selection based on criteria such as activity type, age, and gender. The dataset is rigorously benchmarked with different classifiers and feature vectors, revealing that distinguishing falls from ADLs is relatively straightforward, but identifying specific fall types presents challenges due to similar acceleration patterns. The inclusion of the same subject in training and test data enhances classifier performance, emphasizing individual variability in activity execution.

Inference :The article presents a valuable dataset for human activity recognition and fall detection, offering detailed information for sample selection based on various criteria. The dataset's thorough benchmarking underscores the feasibility of distinguishing between falls and daily activities due to distinct acceleration patterns. However, it also highlights the complexity of differentiating between specific fall types, which often exhibit similar acceleration shapes. The presence of the same subject in both training and test datasets enhances classifier performance, emphasizing the importance of individual activity variability in model training and testing.

4)Kongara Deepika, Gopampallikar Vinoda (2023).ReddyHuman Action Recognition Using Difference of Gaussian and Difference of Wavelet

Abstract:This study introduces a novel action descriptor for Human Action Recognition (HAR) using spatial and spectral filters. The proposed descriptor combines Difference of Gaussian (DoG) and Difference of Wavelet (DoW) features and utilizes Linear Discriminant Analysis (LDA) for dimensionality reduction. Testing on Weizmann and UCF 11 datasets through simulations reveals promising results, with an average accuracy of 83.66% for DoG + DoW on Weizmann and 62.52% on UCF 11, demonstrating improvements in recognition accuracy, particularly in the Weizmann dataset.

Inference:This study presents a novel approach to Human Action Recognition (HAR) by introducing a combined action descriptor using spatial and spectral filters, namely, Difference of Gaussian (DoG) and Difference of Wavelet (DoW). Utilizing dimensionality reduction techniques and a nearest neighbor classifier, the proposed method is evaluated on Weizmann and UCF 11 datasets. Results demonstrate significant improvements in recognition accuracy, particularly in the Weizmann dataset, showcasing the potential of this novel approach for HAR.

5)Liu Yun,Ruidi Ma,Hui Li(2021):RGB-D Human Action Recognition of Deep Feature Enhancement and Fusion Using Two-Stream ConvNet

Abstract:This paper introduces an RGB-D action recognition method called SV-GCN, leveraging video and deep skeleton data. It employs a two-stream architecture, comprising the Nonlocal-stgcn (S-Stream) for skeleton data and the Dilated-slow fastnet (V-Stream) for video data. These streams are fused for enhanced action recognition. Experiments on the NTU-RGB+D dataset demonstrate that the proposed method outperforms existing approaches, significantly improving recognition accuracy in both cross-subject (CS) and cross-view (CV) scenarios.

Inference:This paper underscores the significance of action recognition in various fields, driven by the growing use of high-definition video and the accessibility of skeleton data through depth sensors and pose estimation. It introduces a two-stream network framework that leverages both video and skeleton data, significantly improving recognition performance by addressing their individual limitations. The proposed Nonlocal-stgcn and Dilated-slow fastnet enhance skeleton and video data utilization, making action recognition more robust and accurate.

2.2 Patent Search

2.2.1 ENHANCED HUMAN ACTIVITY RECOGNITION(US20230071636)

Inventor: Stefano Paolo RIVOLTA Roberto MURA Michele FERRAINA

The present disclosure is directed to a device with enhanced human activity recognition. The device detects a human activity using one more motion sensors, and enhances the detected human activity depending on whether the device is in an indoor environment or an outdoor environment. The device utilizes one or more electrostatic charge sensors to determine whether the device is in an indoor environment or an outdoor environment. The device may also exclude gyroscope data when performing human activity recognition, and instead utilize electrostatic charge variation data in conjunction with acceleration data to perform human activity recognition.

2.2.2 HUMAN BODY ACTIVITY RECOGNITION METHOD AND DEVICE

Inventor: HUANG ZEFENG,YU YAWEI,WANG TING

The embodiment of the invention discloses a human body activity recognition method and device, and relates to the technical field of data processing. The method comprises the steps of obtaining original activity data, and preprocessing the original activity data to obtain first data; extracting a time domain feature of the first data to obtain a first feature vector; pre-classifying the first databased on the first feature vector and clustering center points of a plurality of activity categories; and obtaining an activity category to which the first data belongs based on a pre-classification result and a pre-classification confidence coefficient. According to the human body activity recognition method provided by the embodiment of the invention, the lightweight human body activity recognition framework is applied, human body activity recognition does not need to be carried out based on a deep neural network model, and the calculation complexity and occupied hardware resources of a human body activity recognition algorithm are reduced, so that the energy consumption and the response time of the human body activity recognition method are reduced.

2.3 Comparison with Existing Systems

Comparison-

Problem Identification and Objectives:

Our Model identifies pedestrian safety as a pressing concern in urban landscapes and proposes HAR techniques as a solution.

Existing systems also focus on addressing pedestrian safety issues through various methods, including sensor-based detection, computer vision, and machine learning algorithms.

Approach:

Our project proposes using video analysis and machine learning algorithms for real-time detection and classification of human activities.

Existing systems may utilize similar approaches, such as video-based analysis, sensor fusion (combining data from multiple sensors), or deep learning models for activity recognition.

Primary Goals:

Our initiative aims to prioritize the safety of vulnerable groups like the elderly and children while developing a user-friendly interface for stakeholders.

Similar goals are observed in existing systems, which often prioritize safety measures for all pedestrians and may also focus on creating accessible interfaces for stakeholders such as traffic authorities, city planners, and law enforcement agencies.

Potential Impact:

Our project emphasizes the importance of proactive safety measures and timely interventions enabled by HAR technology to mitigate risks to pedestrians.

Existing systems also highlight the potential impact of intelligent systems in interpreting pedestrian behaviors and reducing accidents through timely interventions and enhanced safety measures.

Potential and Challenges:

Our Model highlights the immense potential of HAR techniques in mitigating risks in urban pedestrian environments.

Existing systems also recognize the potential of such technologies but may face challenges related to scalability, cost-effectiveness, privacy concerns, and regulatory compliance.

Chapter 3: Requirement Gathering for the Proposed System

3.1 Introduction to requirement gathering

Requirement gathering is a critical phase in the software development life cycle (SDLC) where the needs, objectives, and constraints of a project are identified and documented. It involves systematically collecting and documenting user expectations, functionalities, constraints, and other necessary information to ensure the successful development of a software system or product. The primary goal of requirement gathering is to establish a clear understanding between stakeholders (such as clients, users, developers, and testers) regarding what the software should accomplish and how it should behave.

3.2 Functional Requirements

● Real-Time Recognition: The HAR system must be capable of real-time recognition of pedestrian activities to ensure immediate response to safety-critical situations.
● Multi-Modal Data Integration: The system should seamlessly integrate data from various sources, including video clips, sensor data, and other relevant sources for comprehensive recognition.
● Scalability: Ensure the system can scale to accommodate different urban settings, from small neighborhoods to densely populated city centers.
● Privacy and Security: Implement robust data privacy and security measures to protect the information gathered by the HAR system and ensure compliance with privacy regulations.
● Performance Evaluation: Regularly assess the systems performance through rigorous testing and evaluation, including metrics such as recognition accuracy, response time, and system reliability.

3.3. Non-Functional Requirements

● Performance: The system must respond within milliseconds to ensure real-time recognition, even in high-traffic urban environments.
● Accuracy: Achieve a high recognition accuracy rate, minimizing false positives and false negatives, to enhance pedestrian safety.
● Usability: Design user-friendly interfaces for both end-users (pedestrians and drivers) and system administrators to enhance ease of use.
● Availability: The system should be available 24/7 to provide continuous

pedestrian safety monitoring and response.

3.4. Hardware, Software , Technology and tools utilized

● Processor - i3 & above
● Disk Space - 8GB
● RAM - 8GB & above
● Programming Language - Python
● Library - Tensorflow, Matplotlib, Pandas, Keras
● IDE - Google Colab/ Jupyter Notebook
● Operating system - Windows 10 & above, macOS, Kali Linux

3.5 Constraints

User-Friendly:
User-friendliness is generally considered a desirable attribute for software systems, achieving it can entail navigating various constraints related to complexity, learning curves, resource limitations, technical dependencies, cultural factors, and regulatory requirements.
Environmental Constraints:
Variability in Environmental Conditions: The HAR system may encounter challenges in accurately detecting and classifying human activities under diverse environmental conditions, such as varying lighting levels, weather conditions (e.g., rain, snow), and background clutter (e.g., crowded streets, moving vehicles).
Integration:
Compatibility with Existing Infrastructure: Integration of the HAR system with existing infrastructure (e.g., traffic signals, surveillance cameras, smart city platforms) may be constrained by compatibility issues, interoperability standards, and technical limitations, requiring careful coordination and adaptation.
Scalability:
The scalability of the HAR system to accommodate increasing data volumes, user demands, and operational requirements may be constrained by limitations in computational resources, storage capacity, and network bandwidth, necessitating scalable architectures and optimization techniques.

Chapter 4: Proposed Design

4.1Block diagram of the system

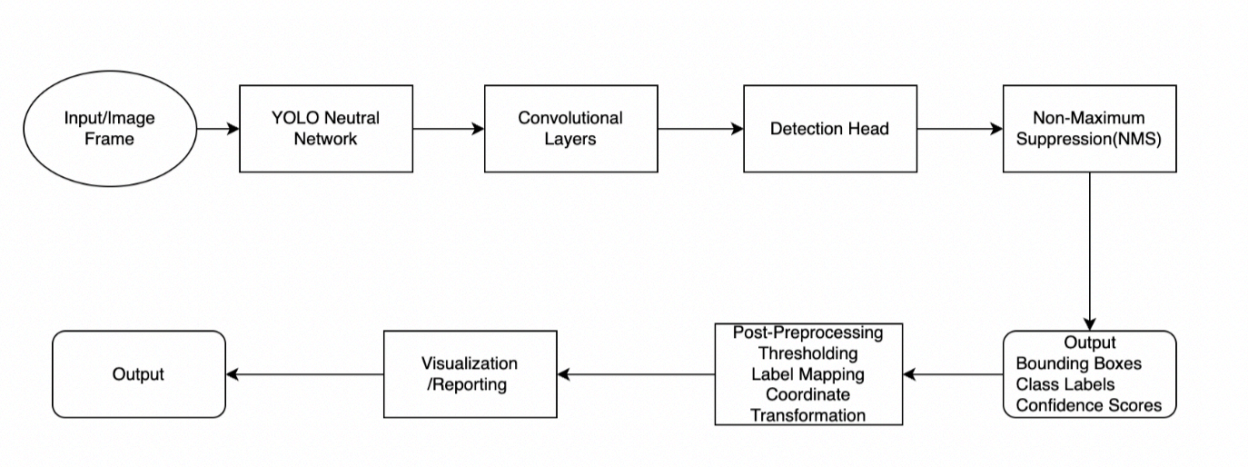


Fig 4.1Block diagram

Input Image/Frame: This is the image or video frame where you want to perform object detection.

YOLO Neural Network:

Input Layer: The input image is passed to the YOLO neural network.

Convolutional Layers: YOLO uses a series of convolutional layers to extract features from the input image. These layers capture patterns and features at different scales.

Detection Head: This is where YOLO predicts bounding boxes, class probabilities, and objectness scores. In YOLO, a grid is applied to the image, and each grid cell predicts multiple bounding boxes and their associated class probabilities and objectness scores.

Non-Maximum Suppression (NMS): After the detection head, a non-maximum suppression algorithm is applied. NMS removes duplicate or highly overlapping bounding boxes, keeping only the most confident one.

Output:

Bounding Boxes: The final set of non-overlapping bounding boxes that represent the detected objects in the image.

Class Labels: Each bounding box is associated with a class label, indicating what type an object has been detected.

Confidence Scores: The confidence score represents how confident the model is that the detected object is of the specified class.

Post-Processing:

Thresholding: You can apply a confidence threshold to filter out low-confidence

detections.
Label Mapping: Convert class indices to human-readable class labels based on a predefined class list.
Coordinate Transformation: Convert the normalized bounding box coordinates to pixel coordinates if necessary.
Visualization/Reporting: The detected objects, along with their class labels and bounding boxes, are often visualized on the original image or reported in a machine-readable format for further analysis.
Output: The final output consists of a list of detected objects, their bounding boxes, class labels, and confidence scores. This information can be used for various applications, such as object tracking, counting, or further decision-making processes.

4.2 Modular design of the system

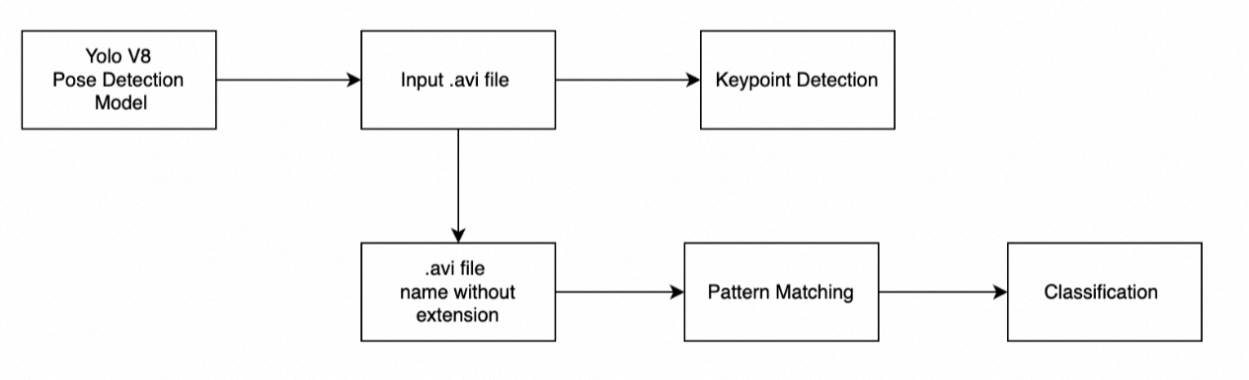


Fig 4.2 Modular Diagram

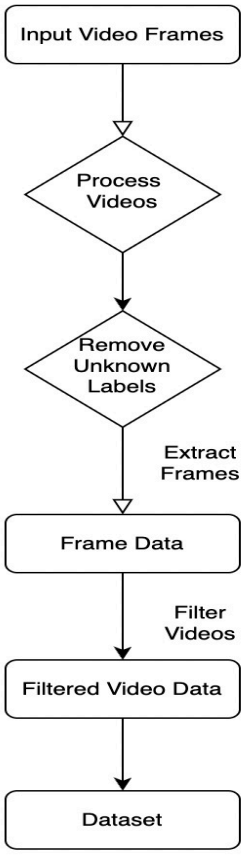
Input (.avi File): This module represents the input stage of the system, where .avi video files containing pedestrian activity data are provided as input. These video files serve as the primary source of data for the subsequent analysis.
YOLOv8 Pose Detection Model: In this module, the YOLOv8 pose detection model is employed to analyze the video frames and detect human poses. YOLOv8 is a deep learning-based object detection model capable of detecting and localizing human body poses within images or video frames.
Pattern Matching: After detecting poses in the video frames, the system performs pattern matching on the file names of the .avi video files. Pattern matching involves comparing the file names against predefined patterns or criteria to extract relevant information. In this context, it may involve extracting specific identifiers or attributes from the file names, such as timestamps or location identifiers.
Classification: Once the poses are detected and the file names are processed through pattern matching, the system proceeds to classify the detected poses based on the extracted information. Classification may involve determining the type of pedestrian

activity observed in the video frames, such as walking, running, standing, or performing specific actions like crossing the road or interacting with objects.

4.3 Detailed Design

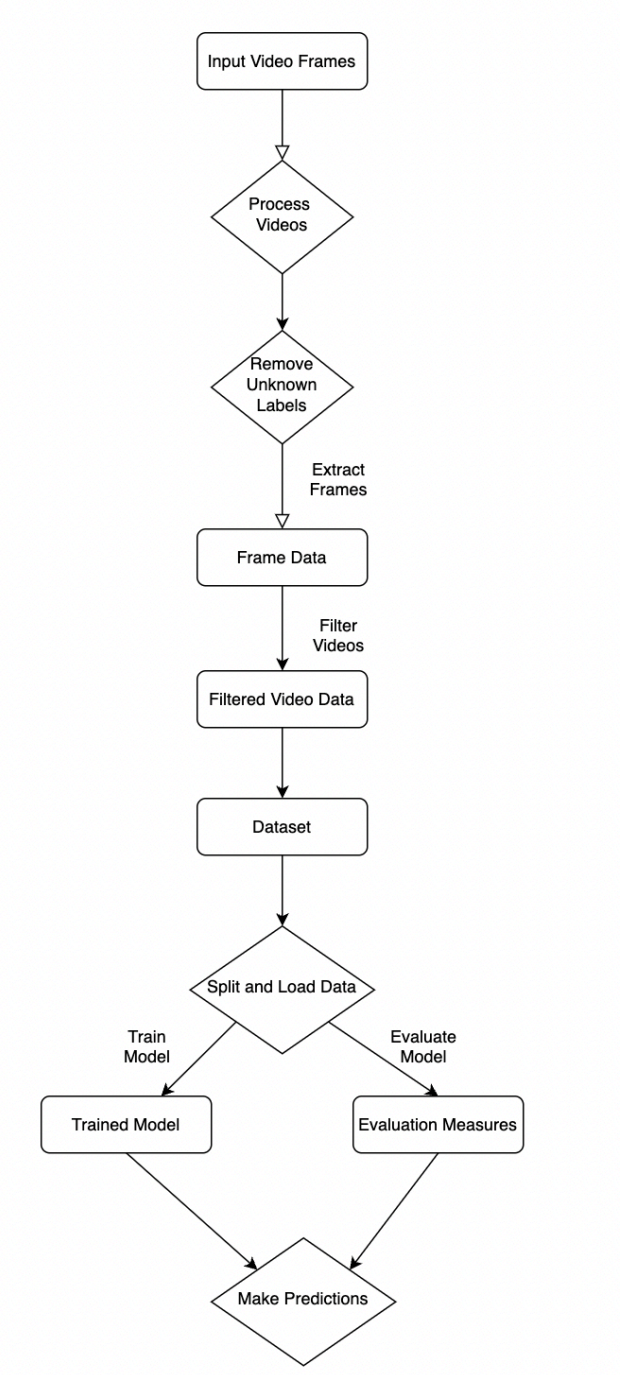
Data flow diagram is a two-dimensional diagram that explains how data is processed and transferred into the system. DFD consists of entities, processes and data stores to show the flow of the data in the system.

4.3.1 DFD Level 0



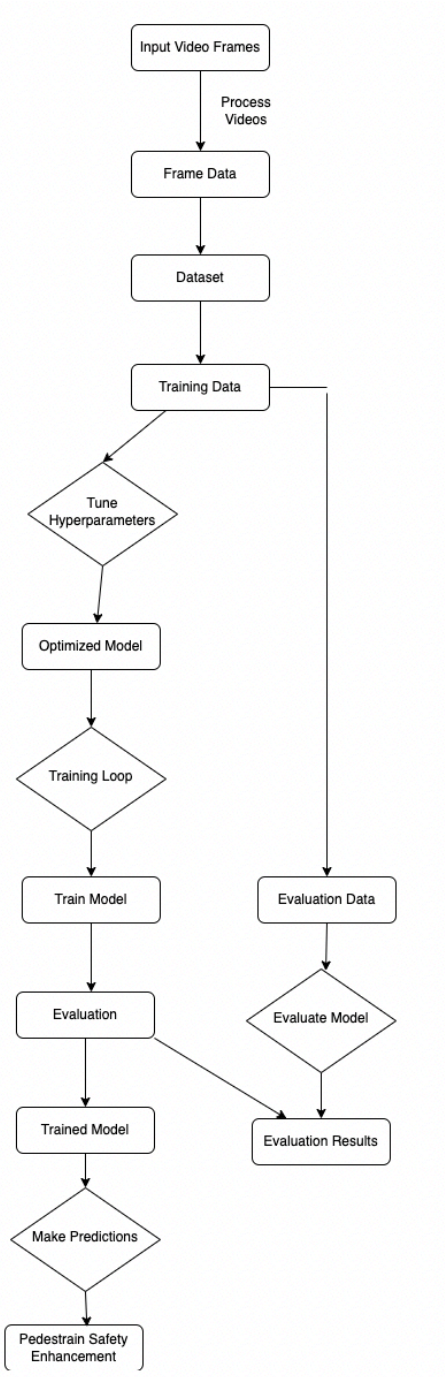
The Level 0 Data Flow Diagram illustrates the initial steps in the processing pipeline for pedestrian safety enhancement. Input video frames are processed to extract relevant information, including removing unknown labels. This refined data forms the basis for generating filtered video data, which ultimately contributes to the creation of a comprehensive dataset for further analysis and model training.

4.3.2 DFD Level 1

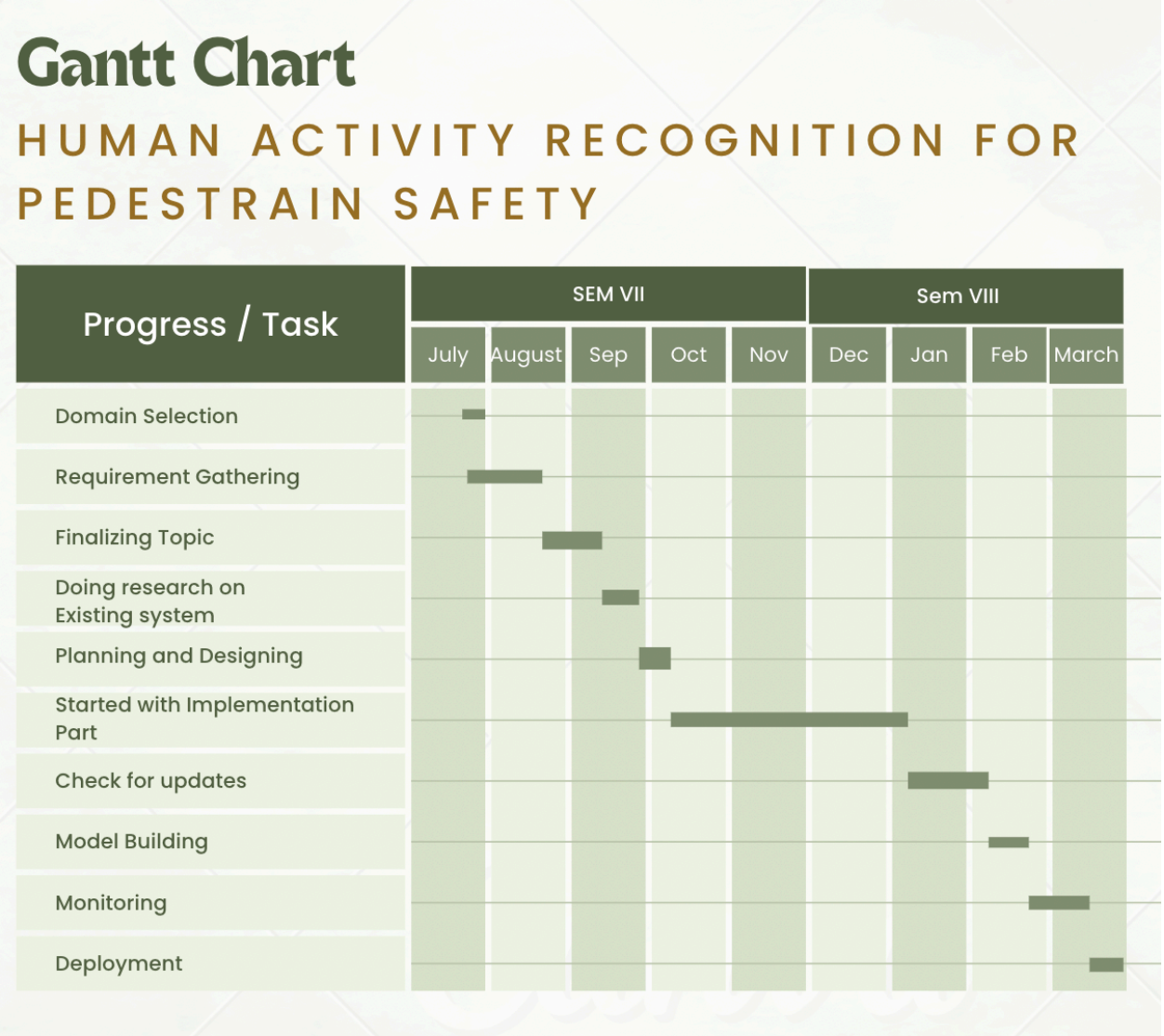


In the Level 1 Data Flow Diagram, video frames undergo processing to eliminate unknown labels, resulting in frame data. This data is filtered to generate a comprehensive dataset, which is then split and loaded for model training. Trained models and evaluation results enable predictions to enhance pedestrian safety.

4.3.3 DFD Level 2



4.4 Project Scheduling & Tracking using Timeline / Gantt Chart



Chapter 5: Implementation of the Proposed System

5.1. Methodology employed for development

YOLO is a popular real-time object detection system that can be applied to various computer vision tasks, including video analysis. The details of the preprocessing and implementation may vary depending on the specific version or implementation you are using, so It'll provide a general overview of how you might preprocess videos using.

● **YOLO:**

Data Collection and Training:
Collect a dataset of video clips or frames with annotated object labels.
This dataset should be diverse and representative of the target Application.

● **Preprocessing Video Data:**

Break down the video into frames or keyframes. You can use video processing libraries like OpenCV to extract individual frames.
Resize or standardize the frame sizes, as YOLO requires a fixed input size for detection.
If your video frames are in RGB format, you can directly use them as input. YOLO can handle RGB image data.

● **Model Architecture:**

Define the YOLO model architecture. YOLOv3 and YOLOv4 are common choices, and each may have a slightly different architecture.
You can use pre-trained YOLO models and fine-tune them on your specific dataset if necessary.

● **Feature Extraction and Fusion:**

YOLO networks typically consist of several convolutional layers to extract features from the input frames.
These features are then fused at different scales to detect objects of varying sizes.

● **Training and Validation:**

Train the YOLO model on your preprocessed video data. This involves optimizing the model parameters to minimize the detection errors.

● Validate the model performance on a separate validation dataset to ensure it generalizes well to new data.

● **Inference and Real-time Detection:**
During inference, you can feed video frames into the trained YOLO model to perform real-time object detection.
The model will provide bounding box coordinates and class predictions for each detected object.

● **Ethical Considerations:**
When using object detection algorithms like YOLO, ethical considerations are essential. Ensure that your system respects privacy and data protection regulations.
Be aware of potential biases in the data and address them in your dataset and model training to avoid discriminatory outcomes.

5.2 Algorithms and flowcharts for the respective modules

● **Input Layer:**
This module represents the input stage of the system, where .avi video files containing pedestrian activity data are provided as input. These video files serve as the primary source of data for the subsequent analysis.

● **YOLOv8 Model:**
In this module, the YOLOv8 pose detection model is employed to analyze the video frames and detect human poses. YOLOv8 is a deep learning-based object detection model capable of detecting and localizing human body poses within images or video frames.

● **Pattern Matching:**
After detecting poses in the video frames, the system performs pattern matching on the file names of the .avi video files. Pattern matching involves comparing the file names against predefined patterns or criteria to extract relevant information. In this context, it may involve extracting specific identifiers or attributes from the file names, such as timestamps or location identifiers.

● **Label Patterns:**
There's a dictionary called label_patterns that maps certain patterns in video names to human-readable labels.For example, if a video name contains "A102," it might be labeled as "side kick." Processing Videos and Labeling:

The program goes through each video file one by one. It extracts the video file's name (without the file extension) to use as a unique identifier. It looks for patterns in the video name and assigns a label based on the label_patterns dictionary. If no pattern matches, it labels the video as "unknown." It uses the YOLO model to predict and detect objects or actions in the video, saving the annotated video with the detected label in the output folder.

5.3 Datasets source and utilization

The dataset used is “NTU RGB+D 120" Action Recognition Dataset. It consists of 3D skeletal data, encompassing a range of pedestrian movements.
From this dataset, we focus on 9 distinct actions relevant to pedestrian safety. These actions serve as the basis for our model's learning process, allowing it to differentiate between safe and risky pedestrian behaviors.

- Safety Risks and Detected Activities

The solution focuses on detecting specific dangerous activities that pose a threat to pedestrian safety, including:

- Punching a person
- Kicking a person
- Pickpocketing
- Pushing a person
- Hitting a person
- Wielding a knife towards another person
- Knocking over another person
- Shooting at another person with a gun
- Grabbing another person's belongings

Chapter 6: Testing of Proposed System

6.1 Introduction to Testing

Testing is a process of finding bugs or errors in a software product that is done manually by the tester. Debugging is a process of fixing the bugs found in the testing phase. In this chapter we performed Quality assurance (QA) to ensure that our application is stable in different networks and user information is secured. To ensure comprehensive Quality Assurance testing of our application, we prepared different test cases that address all aspects of application testing. We make sure that our application employs consistent fonts, style treatments, color scheme, padding between data, icon design, and navigation. We have used different testing and debugging approaches to test our application.

6.2 Types of Tests Considered

Unit Testing: Unit testing is a software testing technique where individual units or components of a system are tested in isolation to ensure they function correctly according to specifications, typically performed by developers during the development process.

Black Box Testing: Black box testing is a software testing method where the internal structure, design, or implementation of the system under test is not known to the tester, focusing solely on validating the system's functionality based on its inputs and outputs.

White Box Testing: White box testing is a software testing approach where the internal structure, logic, and code of the system under test are examined, allowing testers to design test cases based on the understanding of the internal workings of the software.

6.3 Various Test Case Scenario Considered

There are various types of tests that can be considered for software testing. In the context of our grievance monitoring and response system, we will focus on the following types of tests:

- **Unit Testing:**

Unit testing for the Human Activity Recognition (HAR) project involves testing individual units or components of the software system to ensure they function correctly in isolation. Here's an outline of unit testing scenarios for various components of the HAR project:

Data Collection Module:
Test Case 1: Verify that the data collection module can establish connections with cameras and sensors.
Test Case 2: Ensure that the module can capture and store data streams from multiple sources.
Test Case 3: Validate the synchronization of data streams to ensure temporal alignment.

Preprocessing Module:
Test Case 4: Test data preprocessing algorithms for noise reduction and normalization.
Test Case 5: Verify the correct handling of missing or corrupted data.
Test Case 6: Ensure that preprocessing does not introduce distortions or artifacts into the data.

Machine Learning Model:
Test Case 7: Train the machine learning model with synthetic or simulated data.
Test Case 8: Validate the model's accuracy, precision, and recall on test datasets.
Test Case 9: Test the robustness of the model to variations in input data and environmental conditions.

● **White Box Testing:**
White-box testing, also known as structural testing or glass-box testing, involves examining the internal structure of the software system to design test cases that cover specific code paths, branches, and conditions. Here's how white-box testing can be applied to the Human Activity Recognition (HAR) project:

Data Collection Module:
Test Case 1: Test the initialization of data collection objects and ensure that connections to cameras and sensors are established successfully.
Test Case 2: Validate error handling mechanisms for data collection failures, such as device disconnection or communication errors.
Test Case 3: Exercise boundary cases for data collection buffers to ensure proper handling of overflow and underflow conditions.

Preprocessing Module:
Test Case 4: Test individual preprocessing functions (e.g., noise reduction, normalization) to verify their correctness and effectiveness.

Test Case 5: Verify the implementation of boundary checks and error handling for preprocessing parameters (e.g., filter cutoff frequencies, scaling factors).
Test Case 6: Test the behavior of preprocessing algorithms under extreme conditions (e.g., highly noisy input data, outlier detection).

Machine Learning Model:
Test Case 7: Design test cases to exercise different branches and decision paths within the machine learning model, ensuring comprehensive coverage.
Test Case 8: Validate the correctness of model training procedures and parameter initialization.
Test Case 9: Use code coverage analysis tools to measure the coverage achieved by unit tests for the machine learning model.

● Black-Box Testing:
Black-box testing, also known as functional testing, focuses on testing the software system's functionality without knowledge of its internal code structure. Here's how black-box testing can be applied to the Human Activity Recognition (HAR) project:

User Interface Testing:
Test Case 1: Verify that the user interface elements (buttons, menus, input fields) are displayed correctly and are accessible.
Test Case 2: Test user interaction by simulating various user actions (clicks, keyboard inputs) and verifying the expected response.
Test Case 3: Validate the layout and formatting of the user interface across different screen sizes and resolutions.

Data Collection Testing:
Test Case 4: Verify that data collection starts and stops correctly when initiated by the user or based on predefined triggers.
Test Case 5: Test the system's ability to handle different types of input data (video streams, sensor data) and formats (e.g., MPEG, CSV).
Test Case 6: Validate the accuracy and completeness of collected data by comparing it with ground truth annotations or manual observations.

Preprocessing Testing:
Test Case 7: Test the preprocessing module's ability to clean and normalize input data while preserving relevant information.

Test Case 8: Verify the preprocessing module's handling of edge cases and outliers in the input data.
Test Case 9: Validate the consistency of preprocessing results across different datasets and conditions.

Machine Learning Model Testing:
Test Case 10: Verify that the machine learning model produces accurate predictions for different types of pedestrian activities.
Test Case 11: Test the model's generalization ability by evaluating its performance on unseen test datasets.
Test Case 12: Validate the model's robustness to noise, outliers, and variations in input data.

Chapter 7: Results and Discussion

7.1 Screenshots of User Interface (UI) for the respective module

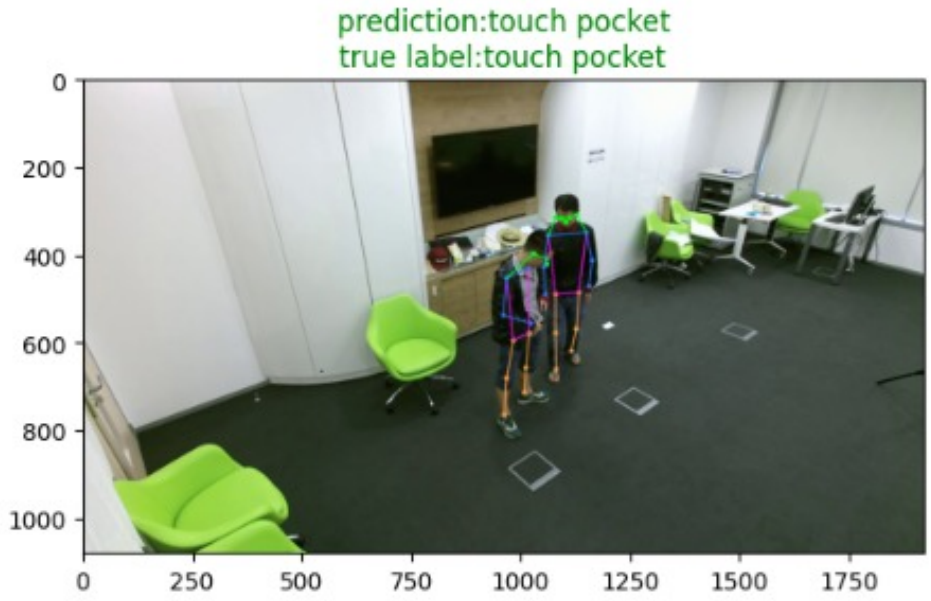


Fig 7.1.1 Pickpocketing

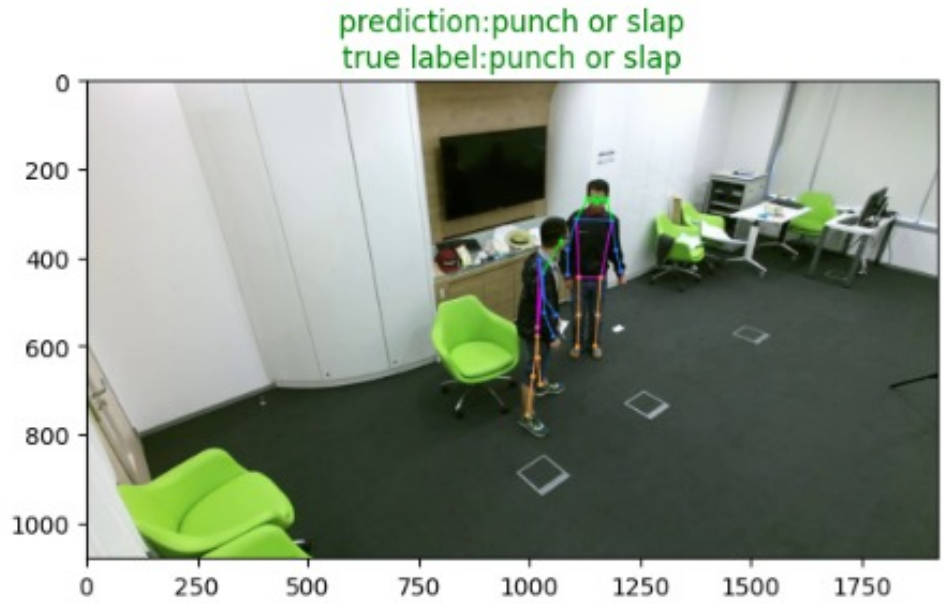


Fig 7.1.2 Punching a person

Fig 7.1.3 Kicking Something

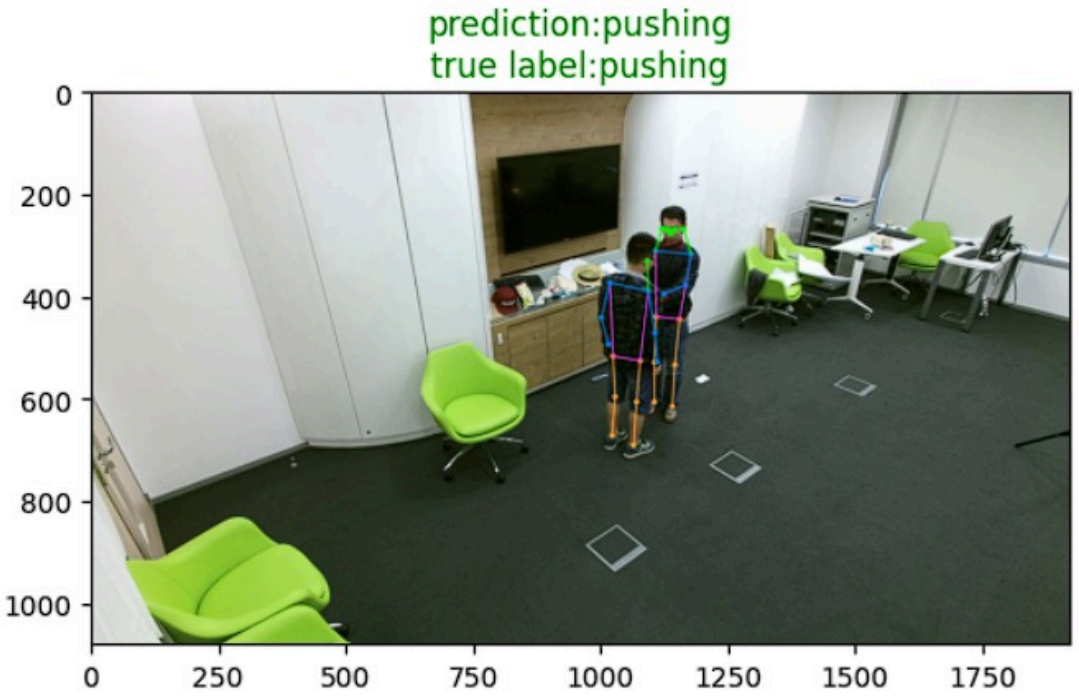


Fig 7.1.4 Pushing a Person

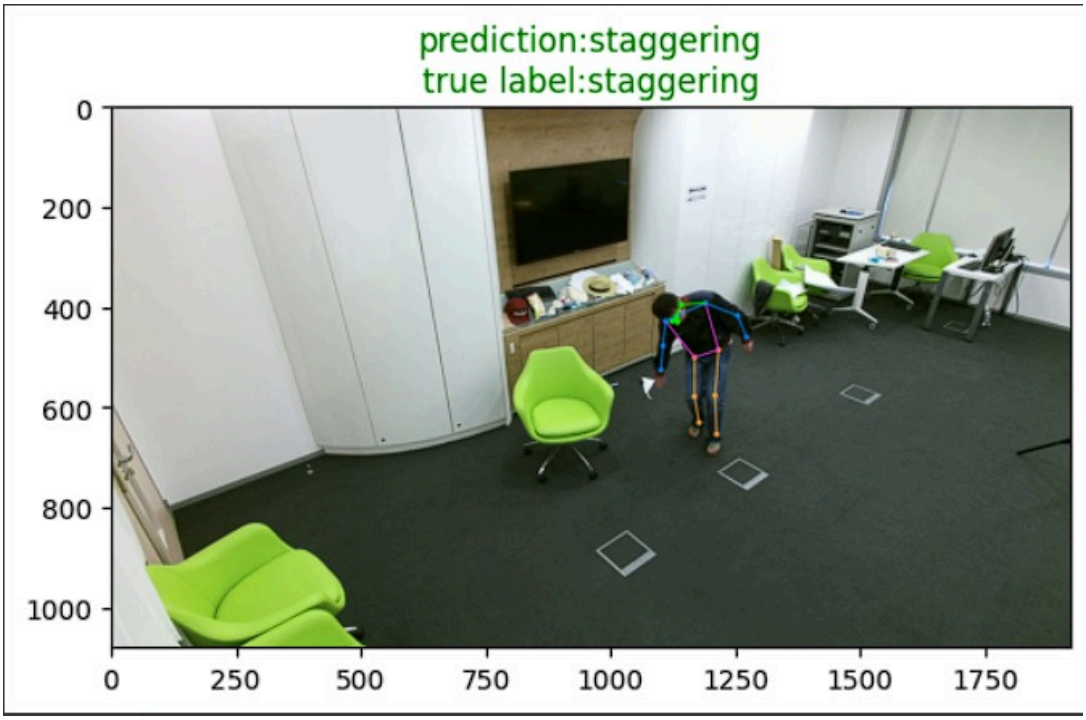


Fig 7.1.5 Staggering

7.2. Performance Evaluation measures

In the realm of machine learning, the confusion matrix emerges as a critical tool for evaluating the effectiveness of classification algorithms. It serves as a visual representation of the model's performance, specifically its ability to accurately distinguish between different categories.

Structure of the Confusion Matrix:

The confusion matrix is a square table organized by the actual and predicted class labels. Each row represents the instances in an actual class, and each column represents the instances predicted by the model. The entries within the table, known as confusion counts, reveal the distribution of predictions:

- **True Positives (TP)**: Correctly predicted positive cases. These instances truly belong to the positive class, and the model accurately identified them.
- **False Positives (FP)**: Incorrectly predicted positive cases. These instances belong to a negative class, but the model mistakenly classified them as positive.
- **True Negatives (TN)**: Correctly predicted negative cases. These instances truly belong to the negative class, and the model correctly identified them.
- **False Negatives (FN)**: Incorrectly predicted negative cases. These instances belong to a positive class, but the model mistakenly classified them as negative

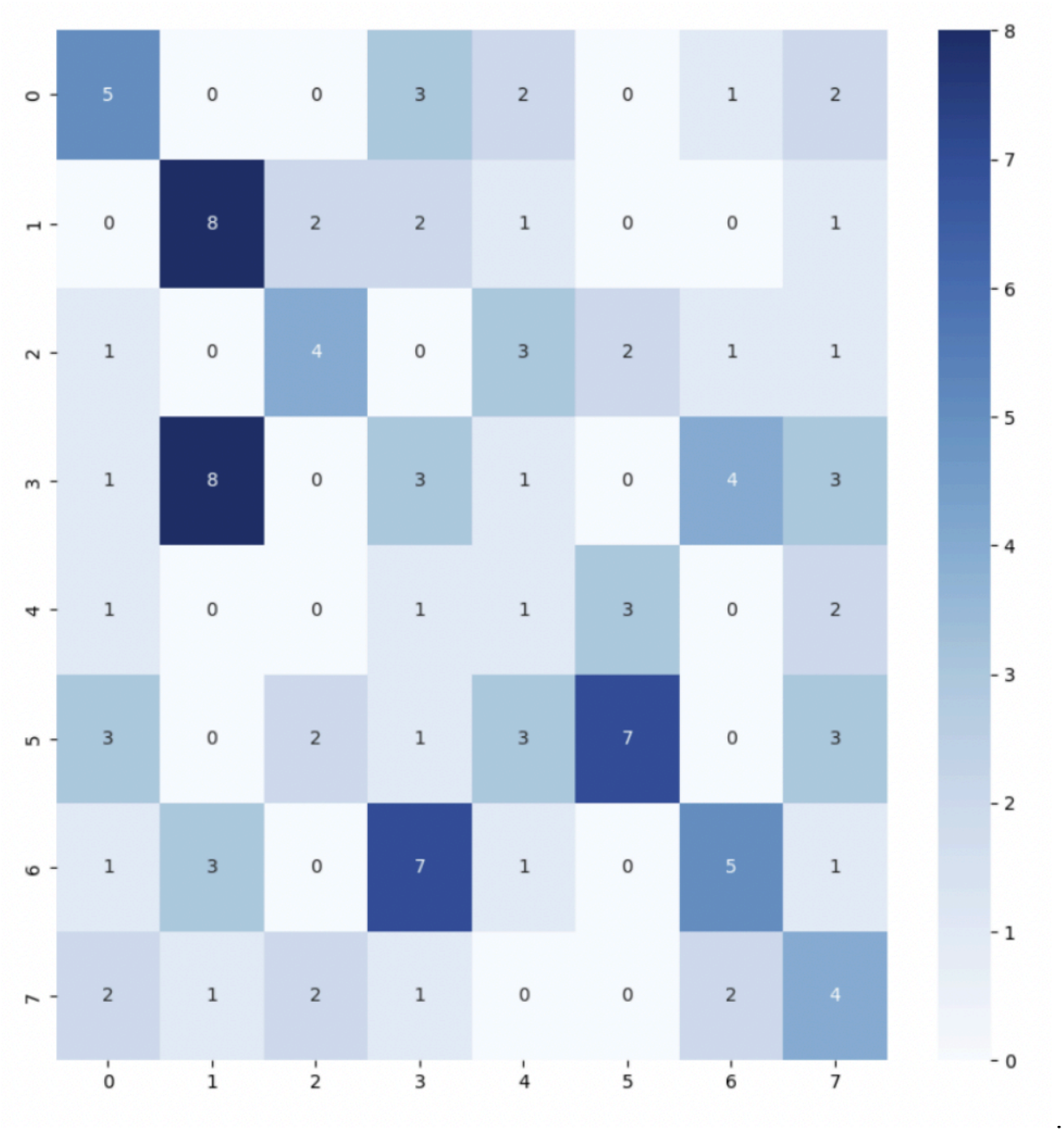


Fig 7.2 Confusion Matrix

Chapter 8: Conclusion

Conclusion

In conclusion, research on human recognition for pedestrian safety, particularly in identifying unusual activities, shows significant promise for enhancing public security. Advanced computer vision and machine learning enable real-time detection of behaviors like punching and pickpocketing, improving overall safety in crowded spaces. While integrating these technologies into surveillance systems offers comprehensive solutions, ethical considerations and privacy concerns must be carefully navigated. This research contributes to safer public environments, showcasing the transformative potential of cutting-edge technology in safeguarding communities. Balancing enhanced security with respect for individual privacy rights is crucial for responsible deployment and widespread societal benefit.

References

- **Jun Lui ,Gang Wang(2019):NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding .IEEE Transactions on Pattern Analysis and Machine Learning
- **Sophie Aubry, Sohaib Laraba*, Joëlle Tilmanne(2019).Action recognition based on 2D skeletons extracted from RGB videos
- **Daniela Micucci,Marco Mobilio(2017).A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones.
- **Kongara Deepika, Gopampallikar Vinoda (2023).ReddyHuman Action Recognition Using Difference of Gaussian and Difference of Wavelet.
- **Liu Yun,Ruidi Ma,Hui Li(2021):RGB-D Human Action Recognition of Deep Feature Enhancement and Fusion Using Two-Stream ConvNet

Human Activity Recognition for Pedestrian Safety

Prof. Vidya Zope
Department of Computer Engineering
VES Institute of Technology(University of Mumbai)
Mumbai, India
vidya.zope@ves.ac.in

Siya Doshi
Department of Computer Engineering
VES Institute of Technology(University of Mumbai)
Mumbai, India
2020.siya.doshi@ves.ac.in

Varun Salvi
Department of Computer Engineering
VES Institute of Technology(University of Mumbai)
Mumbai, India
2020.varun.salvi@ves.ac.in

Sahil Talreja
Department of Computer Engineering
VES Institute of Technology(University of Mumbai)
Mumbai, India
2020.sahil.talreja@ves.ac.in

Roshni Jaisinghani
Department of Computer Engineering
VES Institute of Technology(University of Mumbai)
Mumbai, India
2020.roshni.jaisinghani@ves.ac.in

***Abstract*—The necessity for pedestrian safety is imperative in the hectic settings of urbanisation. With cities becoming more and more bustling, it becomes vital to prioritise pedestrian safety above all else. Pedestrian safety is more than just a preventative measure; it is a fundamental component of urban sustainability.The aim of this project is to utilize the NTU RGB+D 120 dataset and the YOLOv8 algorithm to detect and identify human activities, thereby identifying anomalous pedestrian behavior. By doing so, the project seeks to prevent harm to pedestrians and enhance their safety.**

***Index Terms*—HAR, ML, DL, YOLOv8, NTU-RGB+D 120, Non-local STGCN**

I. Introduction

Human Activity Recognition (HAR) is at the forefront of technological innovation, leveraging advanced methods such as machine learning algorithms and video analysis to detect and categorise human actions. HAR provides detailed insights into the complexities of human activity patterns by scrutinizing and categorising a wide range of human behaviours. HAR's adaptive capabilities extend across a multitude of domains, encompassing healthcare, sports science, and surveillance.

In today's rapidly urbanizing cities, ensuring pedestrian safety has emerged as a critical priority. As urban areas expand and traffic congestion increases, the need to enhance safety measures for pedestrians becomes ever more pressing. To address this challenge, this project adopts a proactive approach by harnessing the power of Human Activity Recognition (HAR) techniques. HAR involves the sophisticated identification and classification of human activities through advanced machine learning algorithms and video analysis. By leveraging HAR technology, real-time monitoring of pedestrian behavior becomes feasible, offering a promising avenue for significantly improving pedestrian safety in urban environments. The primary objective of this initiative is to develop an intuitive interface that caters to the diverse needs of stakeholders, including emergency services, traffic authorities, pedestrians, and drivers. Through strategic integration into existing urban infrastructure, HAR technology can play a pivotal role in fostering safer and more pedestrian-friendly urban environments. By deploying the YOLOv8 algorithm on the NTU RGB+D 120 dataset, this project aims to detect and address anomalous pedestrian behaviors, thereby contributing to the overarching goal of enhancing pedestrian safety in cities.

HAR becomes a powerful ally in the quest for pedestrian safety by coordinating real time observation and behaviour analysis of pedestrians. HAR systems are highly skilled at identifying potentially dangerous pedestrian behaviours, such as jaywalking or careless walking close to roads. Equipped with such knowledge, HAR systems enable prompt alerts and interventions, preventing possible mishaps. Furthermore, HAR's data-rich output helps researchers understand the nuances of pedestrian behaviour and develop clever safety plans. Urban planning frameworks and HAR technology combined allow cities to create pedestrian-friendly areas that are efficient and safe.

II. Background and Related Work

This section discusses the recently built har systems that have been used for their respective case studies using a wide range of data sets.

A. *Human Activity Recognition*

There are several HAR techniques, with each catering to particular applications. These are namely:

1)Sensor-based HAR:
This method records human motion by using information from sensors such as magnetometers, gyroscopes, and accelerometers. It has uses in gesture recognition, healthcare monitoring, and fitness tracking.

2) Vision-based HAR:
This type of HAR uses computer vision techniques to identify human activities by analysing image or video data. It is frequently utilised in applications involving human-computer interaction and surveillance systems.

3)Hybrid HAR:
Increases accuracy and robustness by combining data from vision systems and sensors. This method works well in complicated situations where there are several data modalities available.

4) Deep Learning HAR:
Uses RNNs and CNNs, two deep learning techniques, to automatically extract activity patterns from unprocessed data. It performs extremely well in tasks requiring activity recognition.

5)Context-aware HAR:
Improves activity recognition by taking user and environmental context into account. It modifies activity models for improved performance in dynamic environments by adding contextual information.

B. *Machine Learning and Deep Learning in HAR*

Machine learning (ML) and Deep learning (DL) are indispensable in Human Activity Recognition due to their capability to identify complex and intricate patterns from raw sensor or image data. These algorithms autonomously learn from data, making them well-suited for handling the convoluted nature of human activities. They help the HAR systems to improve their adaptability and enable them to extrapolate according to a variety of human behaviours.

C. *Data used in HAR*

For a comprehensive understanding of human actions, Human Activity Recognition (HAR) makes use of both temporal and spatial data. While spatial data records the actual physical configuration of objects, temporal data records the order and timing of movements over time. Temporal data examines acceleration, velocity, and orientation to detect patterns in behaviour, whereas spatial data focuses on positions and orientations in relation to the environment. The accuracy of HAR is improved by combining temporal and spatial data; this integration is necessary to improve the functionality and efficiency of HAR systems in a variety of applications.

D. *Related Work*

The authors of the study [1], provide an extensive RGB+D human action recognition dataset that includes 8 million frames and over 114,000 video samples from 106 subjects. There are 120 different action classes in this dataset. The effectiveness of current 3D activity analysis techniques is evaluated, and the benefits of deep learning in this situation are illustrated. The authors also discuss the recognition of one-shot 3D activities and provide an efficient Action-Part Semantic Relevance-aware (APSR) framework. The creation of data-intensive learning methods for comprehending human activity is made easier by this dataset.

The paper [2], uses the latest developments in deep learning to present a novel approach for RGB video-based action recognition. Only RGB videos are used for skeletal motion data conversion into 2D images for recognition. Eighteen-joint skeletons are extracted and encoded into RGB channels using OpenPose, a deep neural network. Many encoding strategies were investigated, and ResNet outperformed many state-of-the-art results, 83.3 percent cross-subject, 88.780 percent cross-view accuracy.

In the paper [3], a dataset for fall detection and human activity recognition using smartphone acceleration samples is introduced. It divides activities into 17 classes, including falls and activities of daily living (ADL), using 11,771 samples from 30 subjects. Its meticulous annotation aids in sample selection based on criteria such as activity type, age, and gender. Extensive benchmarking demonstrates that separating falls from ADLs is comparatively simple, but that identifying particular fall types is difficult because of similar acceleration patterns.

Using spectral and spatial filters, the study [4] presents a novel action descriptor for Human Action Recognition (HAR). The suggested descriptor reduces dimensionality by combining Difference of Gaussian (DoG) and Difference of Wavelet (DoW) features and using linear discriminant analysis (LDA). Results from testing on the Weizmann and UCF 11 datasets are encouraging, showing improved recognition accuracy, especially in the Weizmann dataset, with an average accuracy of 83.66 percent for DoG + DoW on Weizmann and 62.52 percent on UCF 11.

The research [5] presents SV-GCN, an RGB-D action recognition technique that makes use of deep skeleton and video data. To improve action recognition, it uses a two-stream architecture that fuses Dilated-slowfastnet (V-Stream) for video data and Nonlocal-stgcn (S-Stream) for skeleton data. Tests conducted on the NTU-RGB+D dataset show

that it outperforms current methods, significantly improving recognition accuracy in cross-subject (CS) and cross-view (CV) scenarios.

III. Methodology Used

A. *Data Analysis*

The dataset we have utilized is NTU RGB+D 120, which is freely available and open-source. This extensive dataset offers an array of resources for RGB+D human action recognition tasks. It adds 60 more action classes to the original NTU RGB+D dataset, making a total of 120 classes. NTU RGB+D 120 offers an extensive collection for training and evaluation with more than 114 thousand video samples and 8 million frames. The NTU RGB+D 120 dataset stands as a cornerstone in the realm of computer vision and action recognition research. It encompasses 3D skeletal data, infrared (IR) videos, RGB videos, and depth map sequences for each sample, captured concurrently on Kinect V2 cameras. The dataset documents a diverse range of human actions, specifically 120 action classes captured within varied indoor environments. We focused on utilizing RGB data in AVI format, characterized by a resolution of 1920x1080 pixels. This dataset serves as a pivotal asset, facilitating advancements in action recognition algorithms and methodologies.

The action classes that we selected for implementing HAR for pedestrian safety are:
1) Punching a person
2) Pushing a person
3) Kicking a person
4) Shooting a person with a gun
5) Weilding a knife towards a person
6) Chest Pain
7) Pickpocketing
8) Staggering

When extracting frames from each video, the model starts from the middle frame to ensure a well-balanced depiction of actions throughout the video sequence. By efficiently capturing contextual information, this method seeks to improve the model's understanding and precise categorization of human activities. The calculation $(\text{total\_frames}//2 - 5)$ determines the starting frame by using the total number of frames $(total\_frames)$ in the video. Deducting 5 from this calculated value ensures that the starting frame is slightly before the exact middle, enabling the model to capture relevant context both before and after the centre point.
Concurrently, feature extraction plays a crucial role in deriving meaningful insights from the extracted frames. The model extracts 10 frames per video and uses those frames to identify key points, which yields an extensive feature set. These features, which total 338 across the dataset, provide an in-depth representation of the actions and movements portrayed in the videos. This large feature set allows the model to learn and distinguish between different human activities, which is beneficial in subsequent classification tasks. The accurate and efficient recognition and classification of human activities is greatly improved by the combination of effective frame extraction and thorough feature representation.

B. *Algorithm*

The algorithm used is YOLOv8. YOLOv8, short for You Only Look Once version 8, is a deep learning-based framework used in Human Activity Recognition. It works by dividing the input image into grid cells and then predicting bounding boxes and class probabilities directly from them. With the help of this method, YOLOv8 can effectively identify and detect human actions in real time. YOLOv8 processes input data and extracts meaningful features related to human activities using a convolutional neural network (CNN) architecture. YOLOv8 gains the ability to precisely detect and categorise a variety of actions through training on large-scale datasets, which enables it to identify a broad range of human activities. Furthermore, YOLOv8 is a well-liked option for HAR applications where real-time processing and high accuracy are crucial because of its benefits like speed, simplicity, and effectiveness.

For applications involving pedestrian safety, the well-known object detection algorithm YOLO (You Only Look Once) can be modified. YOLO is an algorithm that can effectively identify pedestrians in real-time from images or video feeds captured by cameras placed in urban environments. The algorithm is trained on datasets dedicated to pedestrian detection. Because YOLO processes the entire image at once, it is quick and appropriate for applications where it is necessary to quickly identify pedestrians, like autonomous cars or traffic surveillance systems. Furthermore, by accurately detecting pedestrians, YOLO can help reduce potential risks to pedestrians on roads by assisting in the implementation of proactive safety measures and prompt interventions.

In YOLO, a convolutional neural network (CNN) consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers. These layers collectively learn to detect features such as edges, textures, and shapes at different scales and levels of abstraction. This feature map is divided into a grid, with each grid cell predicting bounding boxes and class probabilities. The CNN utilizes convolutional and fully connected layers to make these predictions. Post-processing, like non-maximum suppression, refines the detections. This approach enables real-time object detection by efficiently analyzing the entire image in a single pass.
The HAR model architecture is built with linear layers and consists of 692,232 parameters, allowing the model to learn intricate patterns and relationships in the data. The Adam optimizer with a learning rate of 0.001 is used for training in order to maximise the performance of the model. To track the model's development and identify possible areas for

improvement, its performance is assessed on a different test set at various points during the training process.

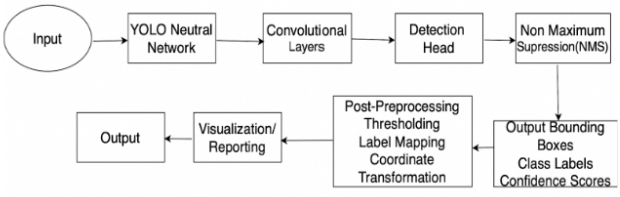


Fig. 1. YOLO Architecture

IV. Experiments and Results

A. *Dataset Visualization*

The model's performance is evaluated using a range of metrics, including accuracy, precision, recall, F1-score, and support. A thorough examination of the model's efficacy for each class provides insightful information. While some classes showed high precision and recall scores, others demonstrated lower performance metrics, suggesting areas for improvement. Class 5 (Pushing a person), for example, had an F1-score of 0.45 based on precision of 0.58 and recall of 0.37. Class 4 (Punching a person), on the other hand, showed lower recall and precision values, with an F1-score of 0.10.

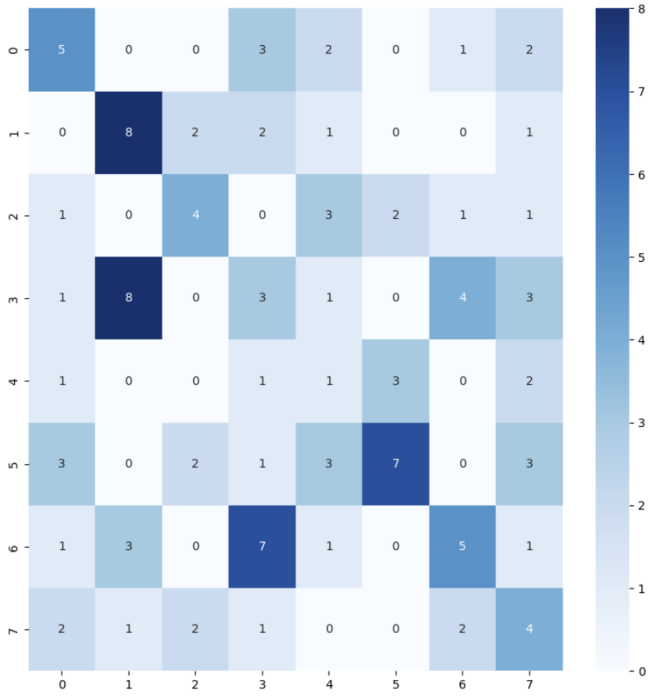


Fig. 2. Confusion Matrix

B. *Results*

In our evaluation, our model rendered precise predictions for a subset of the chosen classes. In particular, we were able to successfully classify the following action classes:
1)Pushing a person
2)Kicking a person
3)Punching a person
4)Staggering
5)Chest Pain
6)PickPocketing

Our experimental results demonstrate that the proposed model accurately classifies human activities. During the training process, the model achieved its highest testing accuracy of 38.41% on the test dataset, demonstrating its potential to extrapolate on new data.

These outcomes underline how well the model recognises and differentiates between these specific human activities. It's important to note, though, that not all classes could be predicted with accuracy; this suggests areas where the model could be strengthened and further refined.

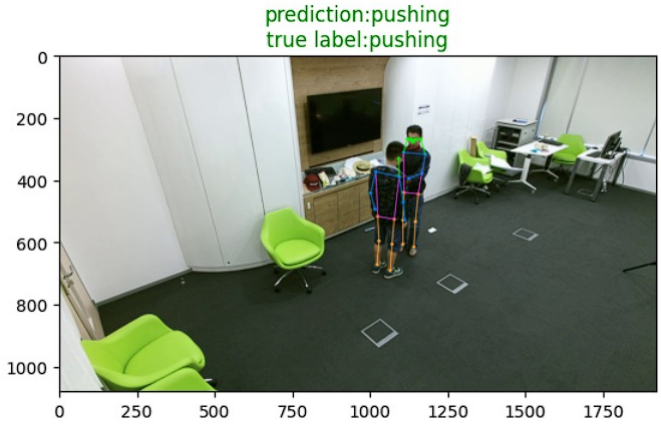


Fig. 3. Pushing a person

Fig. 4. Kicking a person

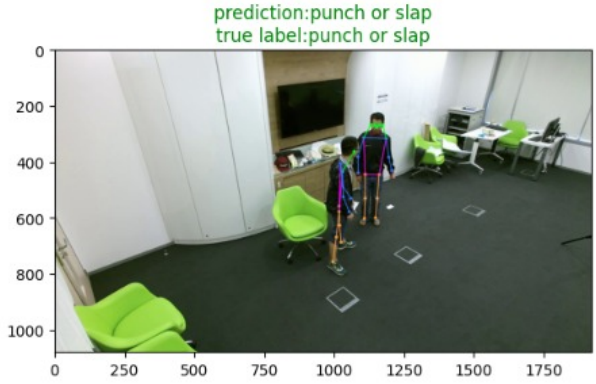


Fig. 5. Punching a person

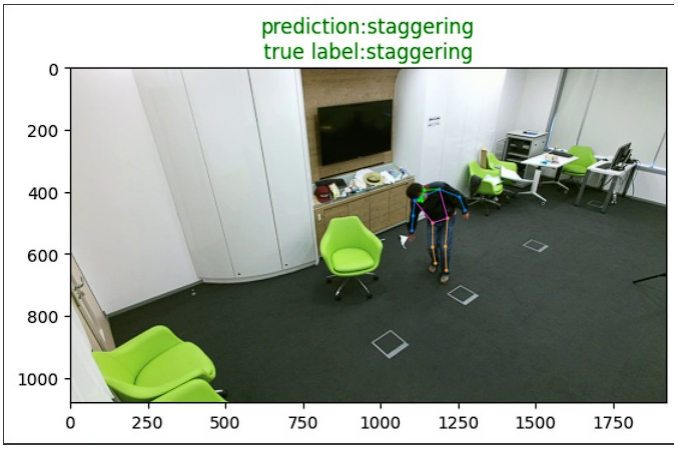


Fig. 6. Staggering

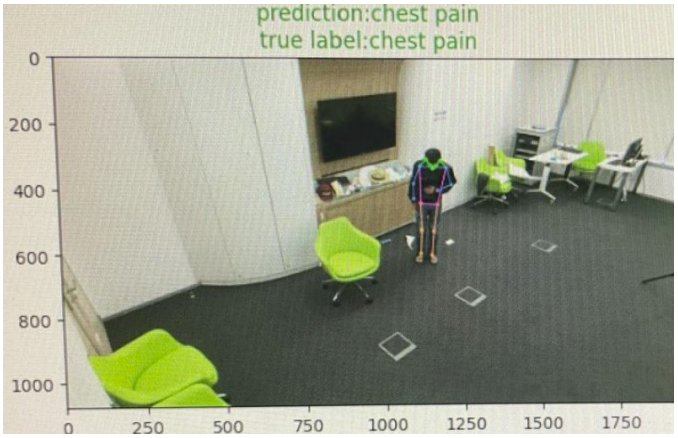


Fig. 7. Chest pain

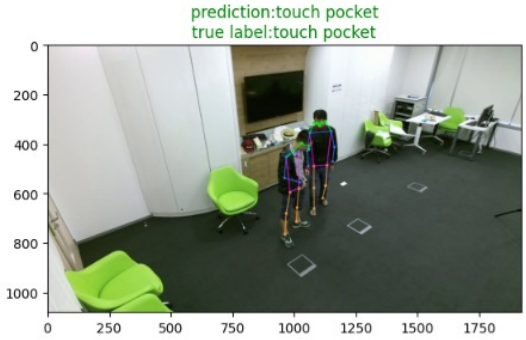


Fig. 8. Pickpocketing

V. Conclusion

In conclusion, our study demonstrates the effectiveness of utilizing YOLOv5 on the NTU RGB+D 120 dataset for real-time detection of anomalous pedestrian behaviors to enhance pedestrian safety. By leveraging RGB video data and state-of-the-art object detection techniques, we achieve reliable detection of abnormal pedestrian actions in real-time scenarios. This approach holds promise for improving pedestrian safety through proactive identification and response to potential hazards on the streets and in public spaces. Future research may focus on further refining the model's accuracy and efficiency, as well as exploring additional datasets and real-world deployment scenarios to validate its effectiveness in diverse environments.

References

[1] Jun Lui ,Gang Wang 2019:NTU RGB+D 120. A Large-Scale Benchmark for 3D Human Activity Understanding .IEEE Transactions on Pattern Analysis and Machine Learning

[2] Sophie Aubry, Sohaib Laraba*, Joëlle Tilmanne 2019. Action recognition based on 2D skeletons extracted from RGB videos

[3] Daniela Micucci, Marco Mobilio 2017. A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones.

[4] Kongara Deepika, Gopampallikar Vinoda Reddy 2023. Human Action Recognition Using Difference of Gaussian and Difference of Wavelet.

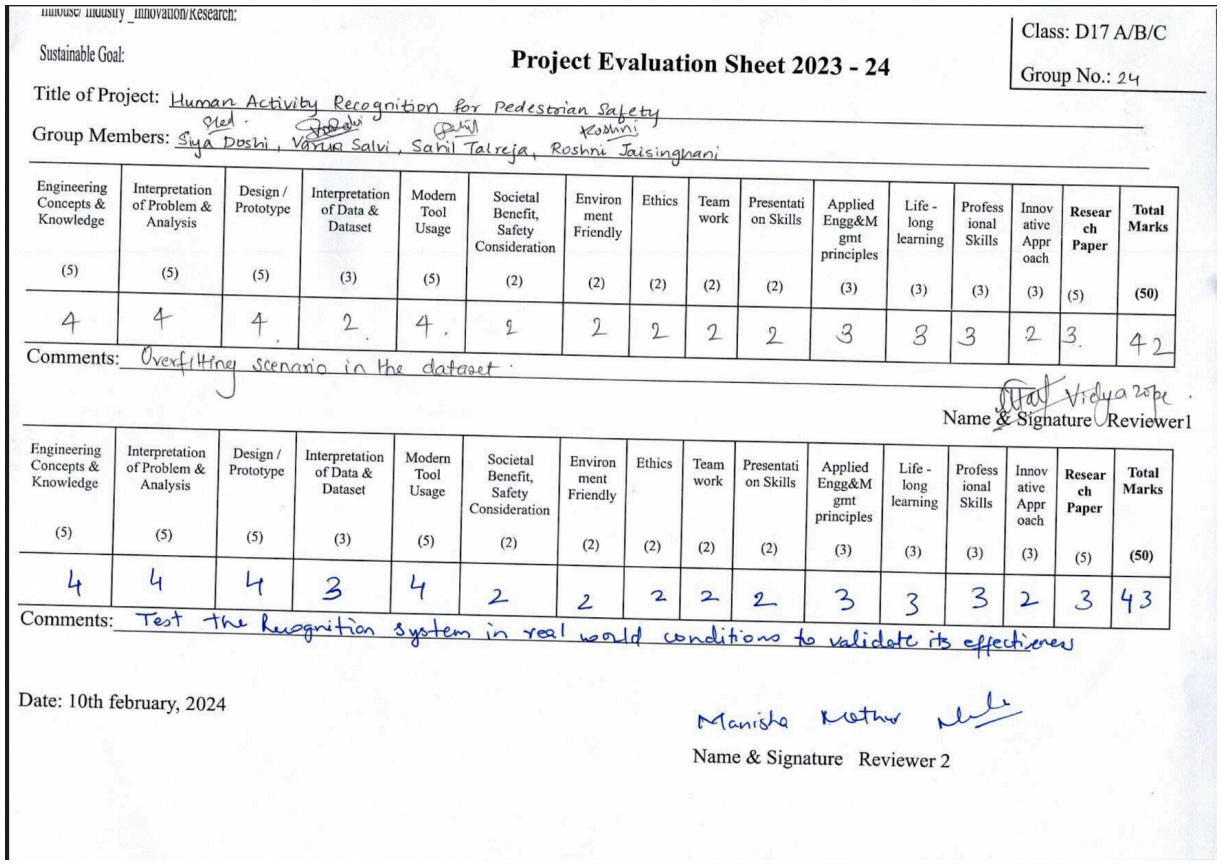
[5] Liu Yun,Ruidi Ma,Hui Li 2021. RGB-D Human Action Recognition of Deep Feature Enhancement and Fusion Using Two-Stream ConvNet

[6] Neil Robertson, Ian Reid 2006. A general method for human activity recognition in video, Computer Vision and Image Understanding.

[7] Vrigkas, M., Nikou, C. and Kakadiaris, I.A., 2015. A review of human activity recognition methods. Frontiers in Robotics and AI.

[8] Aggarwal, J.K. and Xia, L., 2014. Human activity recognition from 3d data: A review. Pattern Recognition Letters.

Project Review sheet 1:



Inhouse/Industry/Innovation/Research:

Class: D17 A/B/C

Sustainable Goal:

Project Evaluation Sheet 2023 - 24

Group No.: 24

Title of Project: Human Activity Recognition for Pedestrian Safety

Group Members: Siya Doshi, Varun Salvi, Sahil Talreja, Roshni Jaisinghani

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life - long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
4	4	4	2	4	2	2	2	2	2	3	3	3	2	3	42

Comments: Overfitting scenario in the dataset.

Name & Signature Reviewer1 Vidya Zope

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life - long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
4	4	4	3	4	2	2	2	2	2	3	3	3	2	3	43

Comments: Test the Recognition system in real world conditions to validate its effectiveness

Date: 10th february, 2024

Manisha Mathur

Name & Signature Reviewer 2

Project Review sheet 2:

Inhouse/ Industry _Innovation/Research:

Class: D17 A/B/C

Sustainable Goal:

Project Evaluation Sheet 2023 - 24

Group No.: 24

Title of Project: Human Activity Recognition for Pedestrian Safety

Group Members: Sahil Tatreja, Varun Salvi, Roshni Jaisinghani, Siya Doshi

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	**Research Paper** (5)	**Total Marks** (50)
4	4	3	2	4	2	2	2	1	2	2	3	3	2	3	39

Comments: Complete Implementation.

Vidya S. Zope

Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	**Research Paper** (5)	**Total Marks** (50)
4	4	3	2	4	2	2	2	1	2	2	3	3	2	3	39

Comments: Put results in Research paper

Date: 9th March, 2024

Mrs. Manisha Mathur 9/3/24

Name & Signature Reviewer 2