JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Implementation Paper: Human Activity Recognition using CNN

[1]Vaishnavi S, [2]Ms. Divyashree S, [3]Shrusti S Rao, [4]Sanjana Shenoy

RNS Institute of Technology, Bangalore, India

Abstract— ***This project centers on the development of a human activity recognition (HAR) system utilizing Convolutional Neural Networks (CNNs) for precise identification of diverse human activities in video footage. The system architecture encompasses three key components: data pre-processing, model selection, and model evaluation. Data pre-processing involves meticulous cleaning and exploratory analysis to optimize model performance. Model selection involves the strategic choice between CNN and recurrent neural network (RNN) models. Model evaluation entails training the selected model on a video dataset and evaluating its performance on an independent test dataset. Harnessing recent advancements in HAR facilitated by Machine Learning (ML) techniques, this project thoroughly explores the potential of CNNs. By transitioning from conventional machine learning approaches to CNNs with automatic feature extraction, the framework aims to improve real-time forecasting of human behaviors and detection of anomalous activities. Targeting domains like senior care and surveillance systems, this implementation contributes to enhancing public safety measures.***

Keywords— ***Human activity recognition (HAR), Convolutional Neural Networks (CNNs), Data pre-processing, Model selection, Model evaluation Machine Learning (ML) techniques, Anomalous activity detection, Surveillance system***

I. INTRODUCTION

Human Activity Recognition (HAR) has become increasingly significant across a wide array of domains including human-machine interaction, healthcare, surveillance, and autonomous driving. There is a growing demand for precise activity detection and pose estimation, catalyzed by the need for efficient time management, contactless interactions amid pandemics, and improved engagement in rehabilitation programs. This survey delves into the aims of training Convolutional Neural Network (CNN) models for automated human activity detection, assessing outcomes, and integrating abnormality detection for real-time surveillance. The focus of the human action recognition system lies in accurately identifying the type of behavior exhibited within a sequence of frames.

II. LITERATURE REVIEW

[1] The article titled "Human Activity Recognition System from Different Poses with CNN" introduces a pioneering strategy for identifying human activities using CCTV footage. It integrates a HAAR Feature-based Classifier for human pose detection with a Convolutional Neural Network (CNN) Classifier for activity recognition. By training on a dataset comprising 5648 images, the system achieves remarkable outcomes, boasting a detection accuracy of 99.86% and a recognition accuracy of 99.82% after 20 epochs. The methodology encompasses three primary phases: human detection/localization, segmentation, and video frame recognition. A comparison with other papers underscores the system's distinctiveness and effectiveness in discerning human activities across various poses. While recognizing limitations such as reliance on a limited activity set, the article advocates for a more diverse dataset to enable comprehensive testing. Overall, the system exhibits promising results in human activity recognition, thereby advancing surveillance and monitoring technologies.

The importance of automating human behavior recognition in real-world scenarios, particularly within video surveillance aimed at bolstering security, is underscored. The paper addresses challenges associated with manually monitoring extensive video data and promotes the integration of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) techniques. It underscores the fusion of visual computing with video monitoring to enhance public safety, introducing advanced Neural Networks such as Convolutional Neural Networks (CNN) and Sequential Neural Networks (RNN) for automating human behavior detection in campus settings using CCTV footage. The proposed system aims to proactively monitor and alert on suspicious events, thereby enhancing both indoor and outdoor security. The training process entails data preparation, model training, and inference, leveraging a detailed framework with a pre-trained VGG-16 model for feature extraction and behavior prediction. The paper encompasses sections on related works, a survey of the proposed method, implementation details, and concludes with reflections on potential avenues for future advancements. This automated system facilitates early detection of anomalous situations, thus improving security measures in diverse environments.

[2] This study introduces a novel hybrid approach for Human Activity Recognition (HAR) by combining Support Vector Machine (SVM) and a 1D Convolutional Neural Network (CNN). Carried out at the University of Missouri, the research addresses the challenge of classifying human activities recorded by smartphone sensors, emphasizing the inherent noise in sensor data and variations in activity signals among individuals. The proposed dual-phase learning strategy initiates with a Random Forest (RF) classifier to distinguish between stationary and mobile activities. SVM is then employed for static activities, while a meticulously designed 1D CNN handles moving activities. The hybrid model achieves an impressive overall precision rate of 97.71% on the UCI-HAR dataset, surpassing or matching state-of-the-art methods. The study underscores the significance of integrating machine learning and deep learning techniques for robust human activity classification, with potential applications in behavior analysis, healthcare, and context-aware computing. This research highlights the importance of automating human behavior recognition in real-world scenarios, particularly within video surveillance systems aimed at enhancing security. It tackles the challenges associated with manually monitoring extensive video data and advocates for the integration of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) techniques. Emphasizing the fusion of computer vision with video surveillance for public safety, the study introduces the application of Deep Neural Networks, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), for automated human conduct detection within campus settings using CCTV camera footage. The proposed system aims to proactively monitor and alert on suspicious events, contributing to both indoor and outdoor security. The training process involves data preparation, model training, and inference, utilizing a comprehensive framework with a pre-trained VGG-16 model for feature extraction and behavior prediction. The paper's structure includes sections on related works, a survey of the proposed method, implementation details, and concludes with reflections on future advancements. This automated system aims to facilitate early detection of unusual situations, thereby bolstering security measures in diverse environments.

[3] This study aims to develop highly precise classification systems for Human Activity Recognition (HAR) using cost-effective hardware technology. Leveraging data from inexpensive sensors like gyroscopes and accelerometers, the research employs an Artificial Neural Network (ANN) in the form of a Deep Stacked Multilayered Perceptron (DS-MLP) for HAR. The DS-MLP comprises an ANN as a meta-learner and five Multilayered Perceptron (MLP) models as base-learners, integrated through a stack

ensemble technique. Evaluation demonstrates an impressive accuracy of 97.3% and 99.4% on heterogeneous datasets. Comparative analyses against existing classifiers and state-of-the-art systems affirm the superior efficacy of the DS-MLP model across various metrics. The study introduces a novel stack ensemble approach for human activity classification, validated on diverse datasets, and exhibits favorable comparison with other prominent HAR systems and machine learning algorithms. The proposed system holds potential applications in healthcare, fitness, and rehabilitation, showcasing the synergy of wearable sensor technology and sophisticated neural network models for precise HAR.

This paper explores the amalgamation of IoT technologies to offer autonomous daily life assistance, particularly for the elderly or individuals with disabilities. The aim is to facilitate extended stays in familiar surroundings through adaptive assistive technologies. The proposed system integrates IoT and assistive technologies to establish a comprehensive assistive system with learning capabilities. It encompasses a smart environment, human activity and health monitoring, an assistive robot, and cloud services, focusing on real-time patient monitoring using IoT for disease prevention and early intervention.The paper presents a prototype human activity and health monitoring system with wireless data transmission for remote analysis. The collected data, including vital signs, facilitates early detection of health issues. Overall, the research envisages enhancing human life quality through continuous monitoring and early detection, laying the groundwork for an integrated assistive system for independent living.

[4] The research outlined in the paper centers on Human Activity Recognition (HAR) using smartphone sensors and selective classifiers. The study utilizes accelerometers within smartphones to detect various physical activities performed by individuals in diverse settings, including walking, jogging, running, stair ascent, and descent. Employing machine learning and deep learning techniques such as Random Forest, Support Vector Machine (SVM), and Convolutional Neural Network (CNN), the authors analyze publicly available datasets to assess the effectiveness of these classifiers, with a particular emphasis on the dominance of deep learning, notably the CNN model, which achieves a recognition accuracy rate of 99%. The proposed method involves data acquisition from a Redmi Note 9 Pro smartphone equipped with a tri-axial accelerometer, followed by preprocessing steps including error handling, label encoding, normalization, and feature scaling. The authors utilize the WISDM dataset, comprising six activity classes, and evaluate the models using confusion matrices, accuracy, and loss curves. The findings highlight CNN's exceptional performance compared to SVM and Random Forest, indicating the potential for accurate human activity recognition through deep learning. The study acknowledges the relevance of HAR in the context of smart devices and Industry 4.0, underscoring the importance of efficient sensor-based methods over vision-based techniques. Despite the promising results, the authors advocate for further exploration and detailed comparative experiments to refine the proposed models.

[5] This article investigates the increasing influx of data propelled by technological progress, particularly in Robotics and the Internet of Things (IoT). With a focus on Human Activity Recognition (HAR), the study meticulously compares the effectiveness of two prominent models—2-D Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM). Employing a consistent dataset obtained from wearable sensors, the models are assessed based on accuracy and confusion matrix analysis. The research highlights the importance of HAR across various domains such as healthcare, smart homes, and fitness tracking. The paper's distinctive contribution lies in its thorough exploration of both CNN and LSTM models, providing insights into their applicability for different scenarios. Furthermore, the study addresses significant challenges in data preprocessing, feature selection, and retraining, paving the way for improved efficiency in HAR. The paper's modular structure encompasses model descriptions, a review of related literature, methodology details, and conclusive results, rendering it a valuable resource for researchers and practitioners in the evolving realm of HAR and IoT applications.

[6] This research article tackles the hurdles in human activity recognition (AR) across life-logging, fitness tracking, and health monitoring domains, where mobile sensing applications are paramount. The authors present an inventive strategy utilizing Convolutional Neural Networks (CNN) to automatically extract discriminative features crucial for precise activity recognition. The approach incorporates a modified weight sharing technique termed partial weight sharing, specifically tailored for accelerometer signals. Unlike conventional AR methods reliant on handcrafted features and domain knowledge, the CNN-based approach enables automatic feature extraction without prior expertise. The benefits lie in CNN's capacity to capture local dependencies and scale invariance in activity signals, crucial for recognizing a wide array of human activities with diverse variations. Experimental results from three publicly available datasets showcase the practicality and superior accuracy of the proposed CNN-based approach compared to existing state-of-the-art methods, highlighting its potential across various application domains, including assembly line tasks, kitchen activities, jogging, and walking.

[7] This paper presents an in-depth examination of recent research papers focusing on sensing technologies employed in Human Activity Recognition (HAR) within the realms of visual computing and human-computer interaction. The study encompasses three prominent sensing technologies: RGB cameras, depth sensors, and wearable devices, assessing their strengths and limitations. The findings indicate that, despite obstacles such as sensor motion and cluttered backgrounds, depth sensors and wearable devices surpass RGB cameras in popularity within current HAR research. The discourse expands to the application of HAR in various scenarios, including surveillance systems for crime prevention in public spaces and healthcare settings such as monitoring elderly individuals in rehabilitation centers and smart homes. The paper concludes by shedding light on the challenges posed by sophisticated sensing technologies, paving the way for further discourse and exploration in this dynamic research domain.

[8] This paper introduces a novel approach to human activity recognition (HAR) utilizing a Convolutional Neural Network (CNN) tailored specifically for tri-axial acceleration signals. The study presents a sizable dataset comprising 31,688 samples across eight typical activities and demonstrates the CNN model's effectiveness, achieving an impressive average accuracy of 93.8% without the need for additional feature extraction techniques. Highlighting the importance of HAR across diverse domains such as healthcare, physical training, and military applications, the paper underscores the advancements in sensor and processor technologies over the past decade, facilitating enhanced precision, compact sensors, and faster processors with reduced power consumption. The proposed CNN-based HAR system operates directly on raw data, eliminating the requirement for a separate feature extraction step and showcasing its potential for swift response. Through comprehensive evaluation and

comparison with existing HAR methodologies, the effectiveness of the proposed deep architecture is substantiated, laying the groundwork for future advancements in this field.

[9] This research addresses a fundamental obstacle in personalized Human Activity Recognition (HAR) using wearable sensors, where the effectiveness of recognition models often diminishes with new users or changes in user physical/behavioral states, requiring retraining with additional labeled data. To surmount this challenge, the paper presents a transfer learning framework employing Convolutional Neural Networks (CNNs) to construct personalized HAR models with minimal user supervision. By capitalizing on representation learning from raw sensor data, particularly with ConvNets, the study aims to augment learning performance without the necessity for expensive data collection and feature engineering efforts, critical in dynamic and uncontrolled wearable environments. The architecture of the developed neural network showcases the potential to efficiently create personalized HAR models, demonstrating the adaptability of deep learning techniques in addressing the challenges linked with contextual changes in wearable sensor applications.

[10] This study investigates the utilization of Convolutional Neural Networks (CNNs) as feature extractors for Human Activity Recognition (HAR) in practical settings, addressing the cold-start problem by evaluating pre-trained CNN models. The research comprises a case study with two primary phases: identifying optimal CNN models for HAR by examining various topologies and parameters, and assessing the pre-trained models on a large-scale real-world dataset, incorporating Inertial Measurement Unit (IMU) and audio-based HAR applications. The achieved balanced accuracy on controlled datasets and real-world datasets highlights the effectiveness of CNNs in feature extraction for HAR, offering contributions in evaluating CNN architectures across sensor modalities and providing insights for overcoming real-world deployment challenges in HAR applications. The paper concludes with an examination of related literature, the proposed case study methodology, experimental particulars, and discussions on results.

III. METHODOLOGY

A. Image pre-processing

Image preprocessing for human activity recognition involves techniques like background subtraction to isolate relevant motion, color space conversion for enhanced feature discrimination, data augmentation for diversifying the training dataset, and histogram equalization for improved contrast and visibility. These steps collectively optimize input data quality, aiding CNN-based methodologies in accurate activity recognition.

B. CNN Architecture

Several CNN architectures have demonstrated efficacy in human activity recognition. Well-known models like AlexNet, VGGNet, and ResNet have been extensively studied. These architectures possess the ability to autonomously learn intricate features extracted from input images, paving the way for precise and resilient activity classification. Additionally, ANN, YOLO, and Human Kinetics are incorporated into the algorithm.

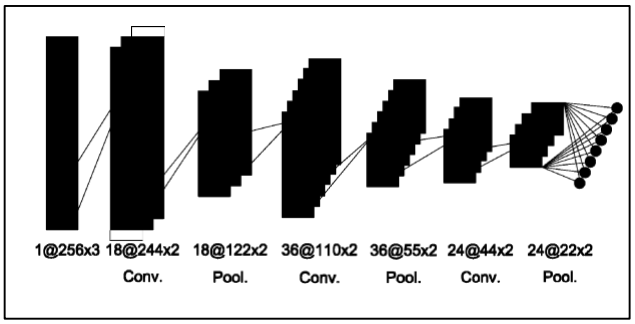

Figure 1: CNN Architecture

C. Feature Extraction in CNN based model

CNNs excel in automated feature extraction, a pivotal element of Human Activity Recognition. These models adeptly discern pertinent features like spatial patterns, textures, and shapes from input images. This section delves into techniques for visualizing and interpreting the features extracted by CNNs, elucidating the underlying mechanisms of the recognition process.

D. Existing System

Recent image-centric systems for Human Activity Identification predominantly rely on traditional machine learning algorithms, although their effectiveness has been surpassed by deep learning methods. The drawbacks of conventional approaches stem from their dependence on intricate handcrafted techniques for extracting features from video frames or images, resulting in a complex process for activity recognition. Utilizing handcrafted methods like Histogram of Gradient (HOG) and Scale Invariant Feature Transform (SIFT) to obtain low-level features may be appropriate for specific datasets. However, the challenge lies in the limited adaptability of handcrafted features to diverse datasets, presenting difficulties in effectively extracting features from new datasets and adjusting manually selected low-level features to different conditions.

E. Motivation

The exploration of human activity analysis stands as a pivotal concern within the computer vision realm, garnering significant attention in recent past. Its applications range from

enhancing intelligent surveillance systems to refining human- computer interactions. While recent methodologies showcase

commendable proficiency in discerning individual actions, the complexity escalates when dealing with collective human activities and their intricate interactions, necessitating a depth beyond individual motion analysis. This research frontier remains formidable owing to the extensive intra-class variation in human activities, arising from visual appearance disparities, subject motion fluctuations, and alterations in viewpoint. The endeavor to comprehend and recognize group dynamics unfolds as a challenging pursuit in the face of these intricacies. Broadly acknowledged as a pervasive research domain, the unobtrusive recording of human activities emerges as imperative, offering indispensable understanding of the rate and duration of functional movements and spinal postures, without impeding the natural range of human motion in a given session.

F. Proposed System

The presented human activity detection application is a comprehensive Python-based solution, seamlessly integrating various libraries such as OpenCV, Pygame, Torch, NumPy, and Tkinter to deliver an intuitive graphical interface. A standout feature is its intrusion detection capability, employing the YOLO (You Only Look Once) model to swiftly detect individuals and objects within a predefined area, triggering alarms upon unauthorized entry or movement, thus fortifying security measures. Additionally, the application incorporates social distance monitoring to ensure adherence to health safety protocols by analyzing camera feeds in real-time, identifying and flagging individuals breaching safe distance rules, thereby aiding in the prevention of viral transmission in crowded spaces, particularly crucial during disease outbreaks. Moreover, the system boasts human activity recognition (HAR) functionalities, facilitating the analysis and classification of diverse human actions from video data. Leveraging pre-trained models on datasets such as Kinetics, it accurately labels various activities depicted in video frames. Furthermore, the application offers live human activity recognition (Live HAR) capabilities, utilizing MediaPipe and machine learning models to detect and classify real-time yoga poses or other activities, providing immediate feedback on pose accuracy, thereby enhancing user engagement and physical activity performance. Overall, the system amalgamates cutting-edge technologies to bolster security, promote health safety, and facilitate activity monitoring across varied environments. The architectural depiction of convolutional neural networks (CNNs) underscores the mechanism's intricacies, delineating the convolutional layer's role in connecting feature map regions, the pooling layer's function in spatial dimension reduction, and the fully connected layers' role in class score computation, all culminating in efficient activity recognition. Key advantages of the proposed system include its cost-effective implementation, user-friendly design, robust recognition of complex activities, seamless integration into daily life monitoring routines, and provision of instant alerts for unusual activity occurrences.

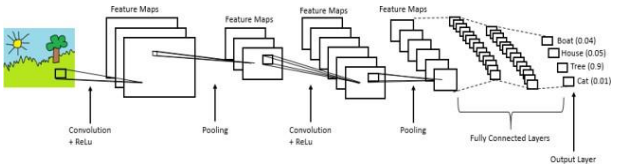

Figure 2: Proposed System

G. Problem Statement

To ascertain an individual's state, it is essential to deploy a network of sensors capable of monitoring various parameters, with a particular focus on physical activities. However, conventional methods present challenges, often requiring expensive setups and fixed infrastructure. Moreover, some approaches necessitate attaching multiple sensors to different body parts, such as wrists, hips, thighs, ankles, and arms, which can be cumbersome and intrusive. Consequently, there is a need to develop a recognition mechanism that aligns with desired features while addressing these limitations. Recent advancements have demonstrated successful implementations of basic human activity recognition; however, a significant drawback persists—many of these achievements rely on sensor units that lack user-friendliness. The field of complex activity recognition remains an ongoing challenge, representing a dynamic area of active research. The ongoing quest is to devise solutions that seamlessly integrate with user needs and expectations without sacrificing usability or comfort.

H. System Design

The process of system design involves the transformation of specified requirements into a tangible implementation, ensuring the incorporation of user-friendly features and efficient functionality. Design specifications delineate the system's features, components, and the appearance of the user interface, thereby facilitating user comprehension and utilization. System requirements outline the necessary conditions for system implementation, providing a comprehensive understanding of project objectives without prescribing specific implementation methodologies. Hardware prerequisites encompass an Intel i5 processor (or a minimum of i3 with 4GB RAM and 80GB HDD), operating at a speed of 2.4GHz, with 4GB RAM, and supporting monitor resolutions of either 1024*768, 1336*768, or 1280*1024. Additionally, a web camera is required. Software prerequisites include Windows 7 or above as the operating system, Google Chrome as the designated web browser, and the installation of Python, Tkinter, OpenCV, and Imutils.

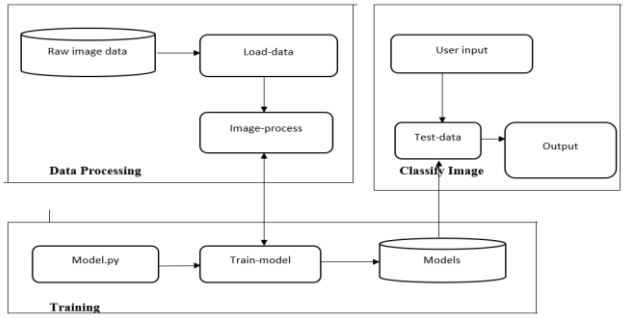

Figure 3: System architecture

IV. ANALYSIS

A. System Architecture

The proposed system architecture depicted in Figure 4 comprises several sequential steps aimed at enabling effective user interaction with the interface. Initially, input data is gathered from diverse sources, including a camera for live video recognition. Subsequently, the input data undergoes processing to extract pertinent information. Advanced AI techniques, encompassing machine learning algorithms and deep learning models, are then employed to recognize human activity and assign appropriate labels. The architecture encompasses interconnected components meticulously designed for efficient video data processing and analysis. At its core lies the video ingestion module, responsible for capturing and streaming video feeds from multiple sources such as webcams or uploaded videos. These feeds are then routed to the preprocessing module, where various techniques are applied to enhance data quality and prepare it for analysis. Post-preprocessing, the video data is channeled to the feature extraction module, tasked with extracting relevant features, potentially utilizing methods like object detection, motion detection, or feature tracking to identify key elements within the video frames. Extracted features are subsequently fed into the activity recognition module, leveraging deep learning models such as CNNs to recognize specific human activities or detect anomalies. Concurrently, the system integrates a rule-based engine or thresholding mechanism to activate alarms or alerts based on predefined rules or thresholds. For example, if the activity recognition module detects unauthorized entry into a restricted area or non-compliance with social distancing guidelines, an alarm is triggered, notifying security personnel or relevant authorities in real-time. Furthermore, the architecture incorporates a user interface module, providing users with a graphical interface to interact with the system. Overall, the system architecture facilitates seamless integration of video analysis techniques, deep learning models, and real-time processing capabilities to deliver advanced solutions for security, safety, and monitoring applications. The figure below (Figure 4) illustrates the general flow of the model.

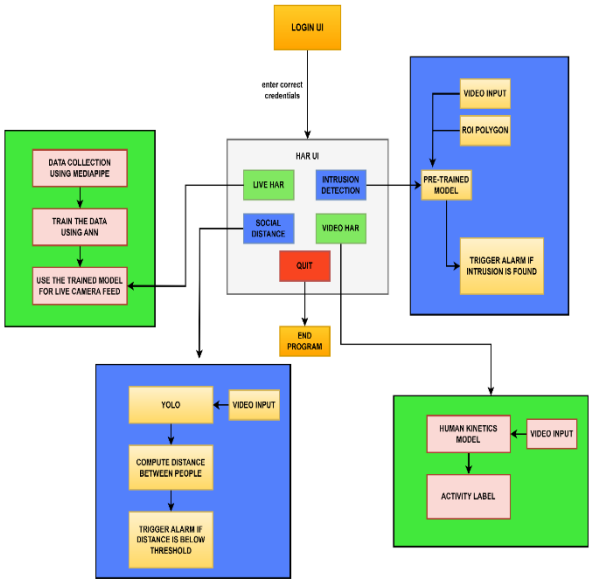

Figure 4: System Architecture

B. *Data Acquisition*

In the load_data.py script, there are functions designed to handle the loading of Raw Image Data and subsequently save the image data as numpy arrays into the specified storage location. Following this, the process_data.py script takes charge of loading the image data from data.npy and proceeds to preprocess the images by resizing or rescaling them, as well as applying various filters and ZCA whitening techniques to enhance features. Throughout the training phase, the processed image data is partitioned into training, validation, and testing datasets, which are then stored accordingly. Additionally, the training process involves the utilization of a load_dataset.py script responsible for loading the relevant data split into a Dataset class. Furthermore, for the purpose of leveraging the trained model to classify gestures, an individual image is loaded and processed

from the file system.

C. Training

The training loop for the model is encapsulated within the train_model.py script. Here, the model undergoes training utilizing hyperparameters sourced from a configuration file, encompassing parameters such as the learning rate, batch size, image filtering methods, and the number of epochs. This configuration, employed during model training, is preserved alongside the model architecture for subsequent evaluation and refinement to enhance results. Within the training loop, both the training and validation datasets are loaded as data loaders, facilitating model training. Post-training, the model is assessed, and the version with the highest validation accuracy is stored for further assessment and application. Upon completion of training, the training and validation error and loss metrics are saved to disk, accompanied by a visualization plotting the error and loss trends throughout the training process.

D. Model Selection: CNNs and ANNs:

At the heart of the system's intelligence lies in the meticulous selection of the model architecture. Recognizing the constraints of conventional machine learning techniques, the system adopts an intricate fusion of Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs). This amalgamation of CNNs, celebrated for their prowess in spatial feature extraction, with ANNs offers a holistic approach to the nuances of human activity recognition. CNNs are harnessed for their adeptness in capturing spatial features from image frames, deciphering patterns in pixel-level data. Concurrently, ANNs play a pivotal role in grasping high-level abstractions and nonlinear correlations within the dataset. The synergy between these architectures ensures a comprehensive grasp of human actions, empowering the system to discern not only the spatial arrangement of activities but also their underlying intricacies and patterns.

E. Real-time activity recognition

The culmination of the system's training endeavors materializes in real-time activity recognition, a pivotal capability with ramifications for various applications like surveillance and anomaly detection. Equipped with the knowledge gleaned from the training phase, the trained model scrutinizes incoming video data to detect and categorize human activities in real-time. This dynamic recognition process entails the continuous analysis of video frames, with the model issuing prompt predictions and refining its understanding of ongoing actions. The real-time aspect holds particular significance in contexts where timely insights into human behavior are imperative for decision-making. The system's output at this juncture furnishes immediate and actionable intelligence, contributing to bolstered security, heightened situational awareness, and the deterrence of irregular behavior.

F. Classify Image

Once a model undergoes training, it becomes capable of classification and is accessible as a file within the file system. When the user inputs the file path of the action's image or frame, the test_data.py script will transmit this path to process_data.py for loading and preprocessing the file in a manner consistent with the model's training procedure.

G. Model Evaluation

To maintain the enduring accuracy and efficacy of the model, a comprehensive evaluation protocol is implemented. Model evaluation entails rigorous testing on distinct datasets, separate from the training data, to gauge its generalization prowess. A suite of metrics, encompassing precision, recall, F1 score, and confusion matrices, is deployed to quantify the model's performance across diverse activity categories. Ongoing evaluation is crucial for the system to adapt to evolving patterns in human behavior and mitigate potential issues like overfitting. Regular enhancements to the model, informed by the outcomes of the evaluation process, further enhance its predictive capabilities. This iterative evaluation cycle guarantees the dependability and sustained performance of the system, particularly in dynamic settings where human activities may evolve over time.

H. Graphical user interface

The system integrates a user-friendly graphical interface (GUI) to facilitate interactive engagement and cooperation between users and the underlying model. Serving as a conduit, the GUI enables users to actively engage in the training process and contribute to the model's adaptability. Through the GUI, users can offer real-time feedback, rectify misclassifications, and provide annotations, all of which directly influence the model's learning trajectory. The interactive dimension introduced by the GUI not only amplifies the system's adaptability but also fosters synergy between human expertise and machine intelligence. This collaboration is particularly pivotal in refining the system's comprehension of intricate human behaviors that may not be explicitly captured in the training data. Consequently, the GUI emerges as a dynamic tool for iteratively enhancing the system based on user inputs.

I. Implementation Overview

The project implementation involves several essential components aimed at developing a comprehensive application for detecting human activity. Utilizing Python and a range of libraries like OpenCV, Pygame, Torch, NumPy, and Tkinter, the implementation focuses on providing features such as intrusion detection, social distance monitoring, human activity recognition (HAR), and live human activity recognition (Live HAR). Each component plays a crucial role in enhancing the

application's functionality and user experience.

Initially, the intrusion detection feature employs the YOLO (You Only Look Once) model to detect people and objects within a specified area in real-time. YOLO's capability for instant object detection makes it suitable for identifying unauthorized entries or movements within a defined polygonal zone. Upon detection, the system activates an alarm to notify relevant personnel of potential intrusions, thereby bolstering security measures through proactive monitoring and response.

Subsequently, the social distance monitoring feature utilizes camera feeds to analyze the spatial relationships between individuals and ensure compliance with health safety protocols. By identifying instances where individuals violate safe distancing guidelines, the system contributes to maintaining a secure environment, particularly in crowded settings. This feature serves as a valuable tool for enforcing social distancing measures and mitigating the risk of viral transmission in public areas.

The human activity recognition (HAR) component focuses on analyzing video data to classify and label various human activities accurately. Leveraging a pre-trained model on the Kinetics dataset, the system can recognize a wide range of activities from video frames with high precision. This capability enables automatic identification and categorization of human actions, facilitating tasks such as video content analysis, surveillance, and behavior monitoring.

Finally, the live human activity recognition (Live HAR) feature utilizes MediaPipe and machine learning models to detect and classify yoga poses or other activities in real-time. By providing immediate feedback on pose or action accuracy, this component assists users in improving their performance and technique. The real-time nature of Live HAR enhances its usability for applications like fitness training, physical therapy, and interactive gaming.

Overall, the project implementation integrates various technologies and methodologies to create a robust human activity detection application with diverse functionalities. By combining computer vision, machine learning, and graphical user interface development, the application caters to a range of use cases spanning security, safety, and health and wellness domains. The modular design and user-friendly interface contribute to its accessibility and effectiveness across different environments and user demographics.

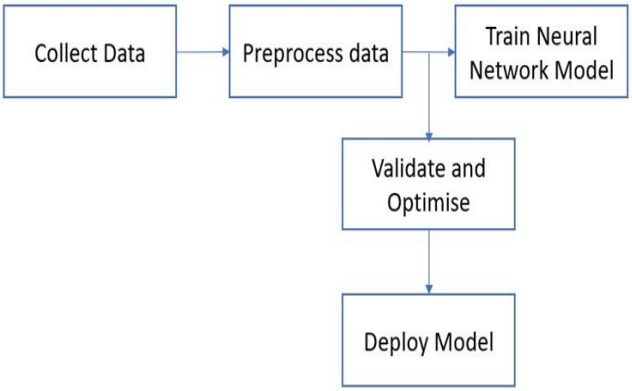

Figure 5: Dataflow Diagram

J. Algorithm

The system initiates by loading necessary libraries for handling video processing, object detection, and graphical user interface operations, setting up a user-friendly environment through Tkinter. Users are prompted to upload videos which can then be processed under various operational modes including intrusion detection, social distance monitoring, and human activity recognition. For intrusion detection, the system employs YOLO for real-time object detection to monitor predefined restricted areas, triggering an audio alert via Pygame if any unauthorized access is detected. Concurrently, the system applies a Human Kinetics model to classify activities in uploaded videos, analyzing and displaying recognized activities on the GUI. In the realm of social distance monitoring, the system utilizes YOLO to identify individuals and calculates the distances between them to ensure compliance with social distancing protocols, issuing visual and audio alerts when violations occur. The live human activity recognition feature leverages MediaPipe for detecting human poses in real-time, classifying these using a pre-trained CNN and updating the GUI accordingly with any notable activities. Interactive GUI components allow users to select specific functionalities such as Live HAR, intrusion detection, and social distance monitoring, facilitating easy user interaction and control. The system is designed to ensure a seamless closure, properly releasing all resources upon termination of the session.

V. RESULTS AND DISCUSSIONS

Human activity recognition represents a crucial aspect across various domains, spanning from healthcare to surveillance and public safety. In our project, we delve into automating activity detection, focusing on user-friendly Convolutional Neural Network (CNN) models. The primary aim is to improve the efficiency of recognizing human activities while integrating real-time alerts for anomalies. The project recognizes the significant potential of deep learning, especially CNNs, in enhancing the

accuracy of activity recognition systems. By embracing this paradigm shift, our initiative addresses the inherent challenges of traditional methods, aiming to make recognition processes more accessible and cost-effective. The integration of CNNs enables a deeper understanding of human activities, overcoming the limitations of conventional techniques. Essentially, our system serves as a cornerstone for public safety, facilitating timely alerts to authorities in cases of restricted area breaches. The convergence of state-of-the-art technology with a dedication to safety positions our project at the forefront of advancements in human activity recognition.

A. User Interface

Upon initial interaction, users are presented with a login interface, prompting them to input their username and password for authentication. This login page acts as the primary entry point for users to access the system, where they verify their credentials to gain entry. It prominently features fields for users to input their username and password. Upon validation of the provided credentials, users are granted access; otherwise, an error message indicating invalid username or password is displayed. Once successfully logged in, users are greeted with a user-friendly menu containing four buttons, each corresponding to a project objective: Intrusion Detection, Human Activity Recognition in Uploaded Videos, Social Distance Monitoring in Videos, and Live Human Activity Recognition.

Figure 6: Login page

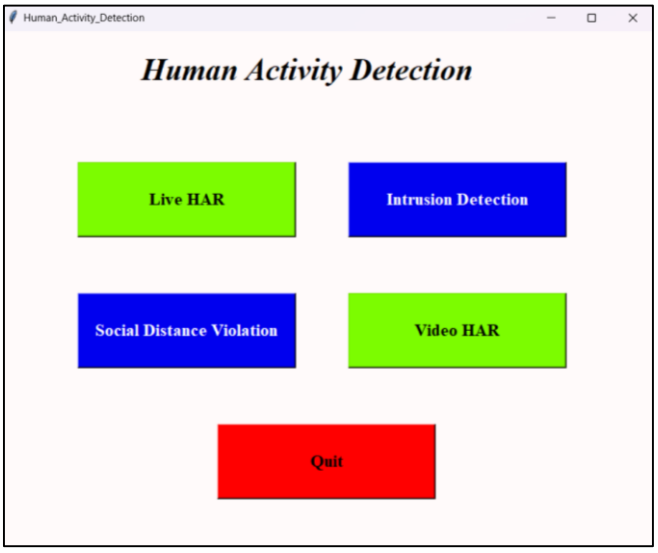

Figure 7: Main menu

B. Live HAR

The live Human Activity Recognition (HAR) system operates through a structured sequence of procedures. Initially, it acquires video data and enhances its quality through various adjustments such as frame resizing and color correction. Subsequently, the system employs a trained CNN model to analyze the live video stream in real-time. This CNN model has been trained extensively to discern diverse human activities, such as walking or waving, through extensive exposure to labeled examples.

During the live video analysis, the CNN scrutinizes each frame to identify ongoing activities. It achieves this by scrutinizing learned patterns and features derived from the training dataset. For instance, it may detect specific movements or shapes indicative of activities like walking or jumping.

Upon activity recognition, such as identifying walking or waving, the system promptly displays this information in real-time. Consequently, when someone initiates walking or waving in front of the camera, the system instantaneously recognizes and

displays the activity. This real-time analysis proves valuable in scenarios like security surveillance, where immediate detection of unusual events is crucial, or interactive systems requiring prompt responses to human actions.

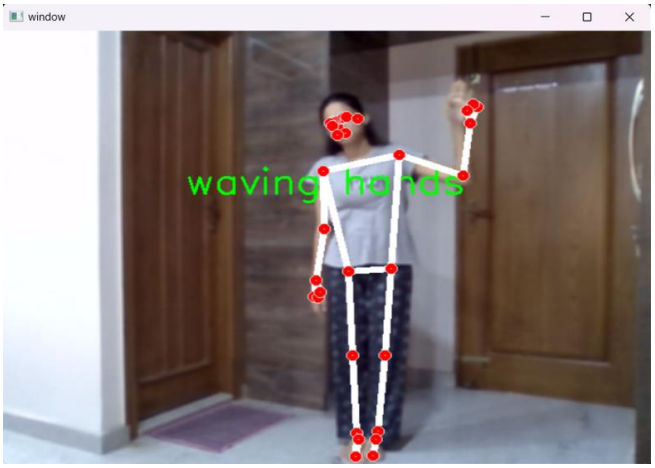

Figure 8: Live HAR Output - Waving

C. Intrusion Detection

The functionality of the intrusion detection system operates through several sequential stages. Initially, the user selects a video file via the application's interface, accommodating various video formats to cater to diverse scenarios, such as surveillance camera monitoring. The visual representation below demonstrates the video playback process within the system.

Subsequently, the user delineates a region on the video to designate the area of interest. This is accomplished by marking points on the video interface to form a shape, such as a rectangle or circle, establishing a virtual perimeter for surveillance.

During video playback, the system employs an intelligent algorithm known as YOLO to analyze each frame adeptly. Renowned for its object detection capabilities, YOLO swiftly identifies objects, such as individuals, and outlines them with bounding boxes directly on the video feed, indicating their locations.

Continuously monitoring the designated area, the system triggers an alarm upon detecting any intrusion into the specified zone. This alarm serves as a clear indication of unauthorized entry, prompting immediate action by alerting the user to the security breach.

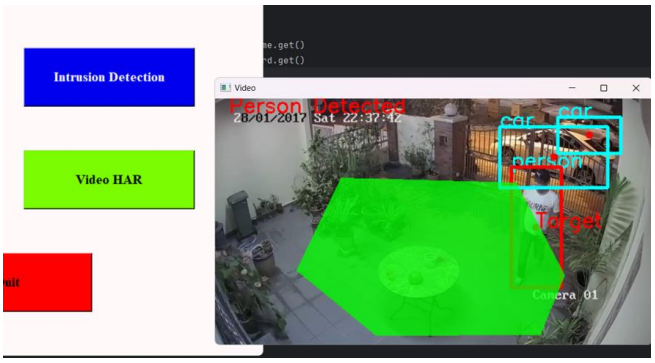

Figure 9: Intrusion Detection Output – Target Spotted

D. Social Distance Violation

The Social Distance Monitoring system is designed to observe video streams, ensuring compliance with safe distancing guidelines among individuals. By analyzing the bounding boxes it generates around people, the system evaluates the distance between them. If the distance falls below the prescribed threshold, indicating unsafe proximity, the system highlights it as a violation by displaying a red box. Conversely, if the distance is adequate, it is denoted by a green box. This intuitive visual representation enables users to promptly discern adherence to safety protocols. Overall, the system effectively oversees social distancing practices, serving as a gentle reminder for individuals to adhere to regulations for the collective well-being, while minimizing unnecessary alarms.

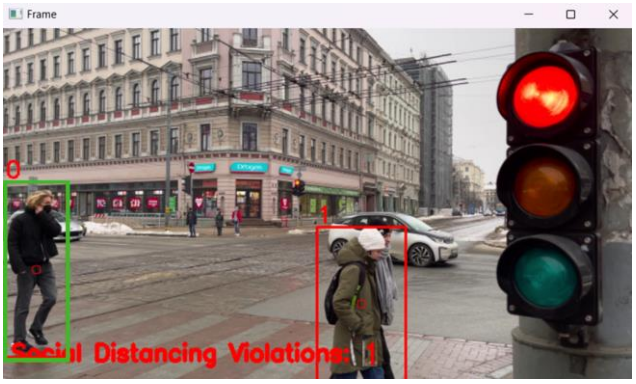

Figure 10: Social Distancing Output – A Single Violation

E. Video HAR

The Human Activity Recognition (HAR) system utilizes Convolutional Neural Networks (CNNs) to analyze uploaded videos and label each frame with the corresponding action observed, such as "running" or "walking." This enables viewers to understand the depicted activities without triggering alarms.

Figure 11: Video HAR Output – Playing Guitar

VI. APPLICATION OF CNN IN HAR

The utilization of CNNs in activity monitoring spans various sectors, including healthcare, sports analytics, and surveillance systems, where CNN models are instrumental in classifying and analyzing human activities. Real-world deployments are examined to illustrate the impact of CNNs on enhancing the precision of activity recognition in different contexts. Apart from healthcare, sports analytics, and surveillance systems, Convolutional Neural Networks (CNNs) also play a crucial role in Human Activity Recognition (HAR) within other domains. One noteworthy domain is smart homes, where CNNs can be applied to identify and adjust to residents' activities, facilitating home automation and bolstering security measures. Moreover, in industrial environments, CNNs contribute to heightening workplace safety by monitoring and categorizing human activities to avert accidents. In the realm of human-computer interaction, CNNs are harnessed to refine gesture recognition systems, enabling more intuitive and responsive interfaces. These varied applications underscore the adaptability and efficacy of CNNs in advancing HAR across diverse fields, illustrating their potential for customized and inventive solutions.

VII. CHALLENGES AND OPEN PROBLEMS

While Convolutional Neural Networks (CNNs) have demonstrated impressive results, several challenges remain in the field of activity monitoring. This discussion highlights issues such as variability in data, the interpretability of models, and constraints in real-time processing. Overcoming these challenges is essential to enhance the precision and efficiency of CNN-driven recognition systems. Beyond these obstacles, the field of Human Activity Recognition (HAR) also faces several critical and unresolved issues. A significant concern is data privacy and security, especially given the sensitive nature of personal activity data. Ensuring that monitoring systems are secure against adversarial threats and unauthorized breaches is key for their broader acceptance. Additionally, deploying HAR in real-world settings often means contending with complex and ever-changing environments, which brings challenges in managing noisy and incomplete data from sensors. Furthermore, the need to scale and adapt HAR models to suit various global populations and cultural settings represents a considerable ongoing challenge. Tackling these comprehensive issues is vital for the future evolution and practical application of reliable and ethical human activity recognition systems.

REFERENCES

[1] Md. Atikuzzaman, Tarafder Razibur Rahman, Eashita Wazed, Md. Parvez Hossain, and Md. Zahidul Islam "Human Activity Recognition System from Different Poses with CNN" 2020 2nd International Conference on Sustainable Technologies for Industry.

[2] Amrutha C.V, C. Jyotsna, Amudha J. "Deep Learning Approach for Suspicious Activity Detection from Surveillance Video" Proceedings of the Second International Conference on Innovative Mechanisms for Industry Applications (ICIMIA

2020).

[3] Stefan Oniga, József SütĘ "Human activity recognition using neural networks" 2014 15th International Carpathian Control Conference (ICCC).

[4] Piyush Mishra, Sourankana Dey, Suvro Shankar Ghosh, Dibyendu Bikash Seal, Saptarsi Goswami "Human Activity Recognition using Deep Neural Network" 2019 Fifth International Conference on Data Science and Engineering (ICDSE).

[5] Md Maruf Hossain Shuvo, Nafis Ahmed, Koundinya Nouduri, Kannappan Palaniappan "A Hybrid Approach for Human Activity Recognition with Support Vector Machine and 1D Convolutional Neural Network", 2020 IEEE Applied Imagery Pattern Recognition Workshop.

[6] FURQAN RUSTAM, AIJAZ AHMAD RESHI, (Member, IEEE), IMRAN ASHRAF, ARIF MEHMOOD, SALEEM ULLAH, DOST MUHAMMAD KHAN, AND GYU SANG CHO "Sensor-Based Human Activity Recognition Using Deep Stacked Multilayered Perceptron Model", Received November 12, 2020, accepted November 24, 2020, date of publication December 2, 2020, date of current version December 16, 2020.

[7] Mst. Alema Khatun, Mohammad Abu Yousuf"Human Activity Recognition Using Smartphone Sensor Based on Selective Classifiers", 2020 2nd International Conference on Sustainable Technologies for Industry

[8] Lamiyah Khattar, Chinmay Kapoor, Garima Aggarwal "Analysis of Human Activity Recognition using Deep Learning", 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence 2021).

[9] Ming Zeng, Le T. Nguyen, Bo Yu, Ole J. Mengshoel, Jiang Zhu, Pang Wu, Joy Zhang, "Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors", Department of Electrical and Computer Engineering Carnegie Mellon University Moffett Field, CA, USA.

[10] Ong Chin Ann, Lau Bee Theng, "Human Activity Recognition: A Review", 2014 IEEE International Conference on Control System, Computing and Engineering, 28 - 30 November 2014, Penang, Malaysia.