Annexure-III

Student Undertaking of Internship/ OJT Academic Details (only applicable when no Internship/ OJT pathway is present in the program scheme)

Name: VEERI DHEERAJ Reg. No.: 12012614

Program Code and Name: P132-NND:: B.Tech.(CSE- Data Science (ML and AI))

Section No.: K20UP

Name of Company: UpGrad Campus Start Date of Internship/ OJT: 18.01.24

Stipend during Internship/ OJT: NA Package: NA

Academic Requirement during Internship/ OJT (To be filled in consultation with Academic HOD and AOC)

Autumn Term (Term id):______________ No. of course to be studied:______________

Details of courses to be studied:

No. of courses to be waived off:

Details of courses to be waived off:

Requirement of CA:

CA is to be prorated as per the provisions of proration policy: _______

Term paper will be assigned in lieu of CA: _______

Any Other: ____

Spring Term (Term id):______________ No. of course to be studied: ______________

Details of courses to be studied:

No. of courses to be waived off:

Details of courses to be waived off:

Requirement of CA:

CA is to be prorated as per the provisions of proration policy: _______

Term paper will be assigned in lieu of CA: _______

Any Other:____

Name of Academic HOD: Name of AOC:

UID of Academic HOD: UID of AOC:

Signature of Academic HOD: Signature of AOC:

Undertaking by Student:

1. I have been informed and I am aware about the academic requirements that I need to fulfill along with OJT/Full term Internship/Full year internship.

1. I understand that I have to fulfill my professional responsibilities in organization and academic
Requirements like ETE/ETP, Field project, CA etc. simultaneously without seeking any favor
From the university.

2. I will manage my leaves in my organization and will appear for ETE/ETPs as per the Examination schedule of University.

3. I understand that if I will not able to appear for exam (due to any reason) then I will appear for reappear/Backlog as per the provisions and schedule of University.

Date: 09.05.24

Signature of Student:

HUMAN ACTIVITY RECOGNITION USING LSTM & CNN

Project Report submitted in fulfilment of the requirements for the Degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

Veeri Dheeraj - 12012614

Supervisor

Mr. AJAY SHARMA

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab (India)

May, 2024

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the B. Tech Project report proposal entitled **"HUMAN ACTIVITY RECOGNITION",** submitted by Veeri Dheeraj at Lovely Professional University, Phagwara, India is a bonafide record of his / her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor

Ajay Sharma

(Name of Supervisor)

DECLARATION STATEMENT

I hereby declare that the research work reported in the Project report proposal entitled **"HUMAN ACTIVITY RECOGNITION"** in partial fulfilment of the requirement for the award of Degree for Master of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor Mr. Ajay Sharma. I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University's Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this Project report represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my Project report work.

Signature of Candidate

Veeri Dheeraj (12012614)

ABSTRACT

In Human activity identification or action recognition nowadays has begun to give birth to a very important achievement of technology which in turn affects the areas like healthcare, security (surveillance), and human computer interaction (HCI). These sensors through features like pre-treatment, extraction, and ML algorithms play a vital role in the field of human movement interpretation (deciphering).

In the centre of all activity monitoring stands several sensors, which are very different from each other. Wearable sensors in the forms of accelerometers are used to capture real-time body movements as strategically placed ambient sensors record changes in the environment and associate them with these activities. Self-evident computing gadgets make the data environment even more opulent and allow to study human behaviour more completely.

Significant meaning is being created from the original sensor data that's received. Methods, for example, time-frequency features help to identify temporal aspects of movement while frequency domain features might reveal some of the rhythmic movement patterns. Spatial temporal attributes, which include information gathering from various sensors, help to compose a full picture of movements in a space. Data obtained from sensor signals can give us a range of features, which the researchers can use to translate the hidden information underneath the language.

Machine learning algorithms then act on the extracted features using their capabilities. The SVMs and RNNs are the traditional classifiers that embrace complexity and determine an activity while the newer technologies like CNNs recognize patterns that are multi-dimensional and feature-based, thereby enabling intricate activity recognition. undefined the presence of data fluctuations necessitates algorithms with robust characteristics, real-time processing engenders the need of efficient algorithms, and integral question is model interpretability. Finally, the solution to all mentioned challenges is cooperation between the researchers in together fields of signal processing, machine learning, psychology, and human factors engineering.

Implementing the human activity detection presents a vast range of prospects for the future. Multimodal Sensor Integration allows much greater precision; this coupled with the possibility of lightweight algorithms means that the integration of wearables will be much more pervasive. The data of the standardized and the interpretable models should be used to get the unbiased results and build the social trust. As technology continues to advance with sensors growing smarter and machine learning algorithms becoming more sophisticated, the human activity detection will remain changing the way we understand and relate to how our people moves.

Keywords—Human Activity Detection, Long- Short term memory, Computer Vision, Machine Learning

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have helped me with the development of this HUMAN ACTIVITY RECOGNITION. This project would not have been possible without the support and guidance of several individuals and organizations. Firstly, I would like to thank our project supervisor Mr. Ajay Sharma for providing us with valuable insights and guidance throughout the development of this system.

Their continuous support and encouragement have been instrumental in the successful completion of this project. I would also like to extend our gratitude to our colleagues and classmates who have provided us with constructive feedback and suggestions. Their input has helped us to refine our approach and improve the performance of our project.

I am also thankful to the developers of the various open-source libraries and frameworks used in the implementation of this project. These include Python, Pandas, NumPy, Scikit-learn, and nltk among others.

This technology has played a significant role in the accuracy and effectiveness of the LSTM and various other algorithms used in the system.

Finally, I would like to acknowledge the contributions of the online Railway services, which provided the data used in the development of the information of the system. We are grateful for their efforts in maintaining and updating the database, which has been an invaluable resource for this project.

In conclusion, I would like to express our heartfelt appreciation to everyone who has played a role in the development of this ai-powered railway information system: enhancing user experience through voice interaction and chatbot. Your support and contributions have been invaluable, and I am grateful for the opportunity to have worked on this project.

TABLE OF TOPICS

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Human activity detection, a fundamental task in the field of computer vision and artificial intelligence, aims to automatically recognize and classify human actions from video or sensor data. This area of research has gained significant attention due to its wide range of applications across various domains, including surveillance, healthcare, sports analysis, human-computer interaction, and robotics. The ability to automatically detect and understand human activities has numerous practical implications. In surveillance, it enables the automated monitoring of public spaces, enhancing security and safety measures. In healthcare, it facilitates the tracking of patient movements and activities, aiding in rehabilitation. Programs and elderly care. Moreover, in sports analysis, it provides valuable insights into athlete performance and training strategies.

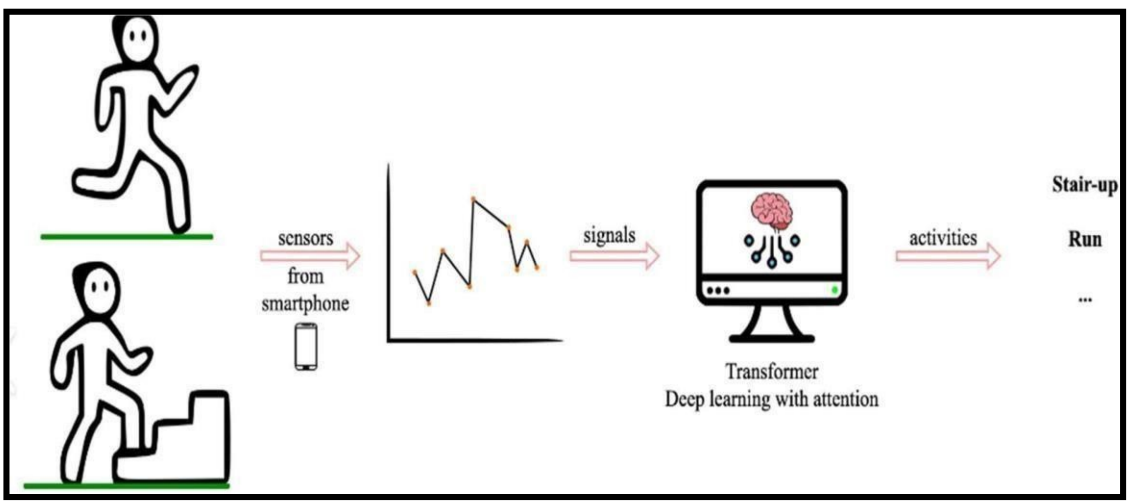

Fig.1.1. Human Activity.

Understanding human behavior has been a longstanding pursuit across multiple disciplines, from psychology to engineering. In recent years, the advent of advanced technologies, particularly in sensor design and machine learning algorithms, has enabled a significant leap

forward in our ability to detect and interpret human activities automatically. Human activity detection, also known as activity recognition, encompasses a range of methodologies aimed at deciphering the myriad gestures, movements, and actions performed by individuals in various contexts. By harnessing data from sensors such as accelerometers, gyroscopes, cameras, and microphones, coupled with sophisticated algorithms, researchers, and practitioners can discern patterns in human behavior with unprecedented accuracy and efficiency. This field has transcended its academic origins to become a cornerstone in the development of intelligent systems and applications. In the realm of smart homes and environments, activity detection facilitates seamless automation of tasks based on occupants' behaviors, enhancing comfort, convenience, and energy efficiency. In healthcare settings, it empowers clinicians with tools for remote patient monitoring, fall detection, and early intervention, thus improving the quality of care and extending independent living for the elderly and individuals with disabilities. Furthermore, in industrial contexts, activity detection contributes to ensuring workplace safety, optimizing production processes, and enhancing human-robot collaboration.

1.1 MOTIVATION

Recognizing human activities through technological means has emerged as a critical area of research, with profound implications across various domains such as healthcare, surveillance, and smart environments.

The ability to automatically detect and classify human actions from sensor data holds promise for enhancing the quality of life and improving efficiency in numerous applications. Traditional methods in human activity recognition often relied on manually engineered features and simplistic learning algorithms, which struggled to capture. the intricate and dynamic nature of human movements. However, the advent of deep learning has ushered in a new era in HAR research, offering the potential for more accurate, robust, and adaptable activity recognition systems. By leveraging the power of neural networks to automatically learn hierarchical representations from raw data, deep learning models have demonstrated unprecedented capabilities in discerning complex patterns and variations in human activities.

Fig. 1.2. Different activities of Human

The significance and applications of human activity detection lies in various applications cross various domains, including healthcare, security, entertainment, and smart environments.

Healthcare Monitoring: Human activity detection can be utilized in healthcare settings for monitoring the daily activities of patients, especially elderly individuals, or those with chronic illnesses. By tracking activities such as sleeping patterns, walking, or eating habits, healthcare providers can assess a patient's health status remotely and detect any anomalies or changes that may indicate health issues or potential risks -

- **Assistive Technologies:** In assistive technologies, such as smart homes or wearable devices, human activity detection can assist individuals with disabilities or elderly populations in maintaining independence and safety. By recognizing activities like getting out of bed, preparing meals, or taking medications, these systems can provide timely assistance or alerts in case of emergencies.

- Security and Surveillance: Human activity detection plays a crucial role in security and surveillance applications, including monitoring public spaces, workplaces, or private properties. By detecting suspicious or abnormal activities, such as trespassing, loitering, or unauthorized access, these systems can enhance security measures and help prevent crimes or accidents.

- Smart Environments: In the context of smart environments, such as smart cities or smart buildings, human activity detection contributes to optimizing resource usage, improving energy efficiency, and enhancing overall user experience. By understanding human behaviours within these environments, automated systems can adjust lighting, heating, or ventilation systems, accordingly, leading to energy savings and environmental sustainability.

- Human-Computer Interaction: Human activity detection enables more intuitive and responsive human-computer interactions, particularly in applications involving gesture recognition, motion tracking, or virtual reality. By accurately capturing and interpreting human movements and gestures, these systems can enhance user interfaces, gaming experiences, and immersive simulations.

- Behavioral Analysis and Research: Human activity detection facilitates behavioural analysis and research in various fields, including psychology, sociology, and anthropology. By studying patterns of human behavior and interactions, researchers can gain insights into social dynamics, cultural practices, and individual preferences, contributing to the development of theories and interventions aimed at improving human well-being and societal outcomes.

1.2 PROBLEM STATEMENT

In the age of technology where gadgets are so integrated in people's lives, the requirement for sensors that recognize smart human behavior has exponentially increased. This requirement originates from a wide variety of realms such as healthcare, sports analytics, security, and human computer interaction.

The basic obstacle is to think through, develop, and improve a complex human activity recognition system that is able to distinguish the unique movements of the human body with a high level of precision and efficiency. The system will utilize the information coming from different sensors including the inertial sensors, cameras and environmental sensors, while overcoming the challenges posed by changes in human behaviour, environmental variations and sensor noise. The ultimate purpose of this project is the design of innovative human activity recognition solution that apart from the precision and real-time implementation also takes dynamic adaptation of the environment and users' preferences into account, thus, improving the experience of users, enabling personalized services and igniting the revolution of human-oriented computing.

1.3 METHDOLOGY

Our methodology encompasses several stages, including data preprocessing, model design, training, and evaluation. We describe the preprocessing steps involved in preparing the UCF50 dataset for training, including video normalization, frame extraction, and annotation. Subsequently, we introduce the architecture of our deep learning models, which combine CNNs and LSTM to capture spatial and temporal features from video sequences effectively. Details regarding model hyperparameters, optimization techniques, and training strategies are provided.

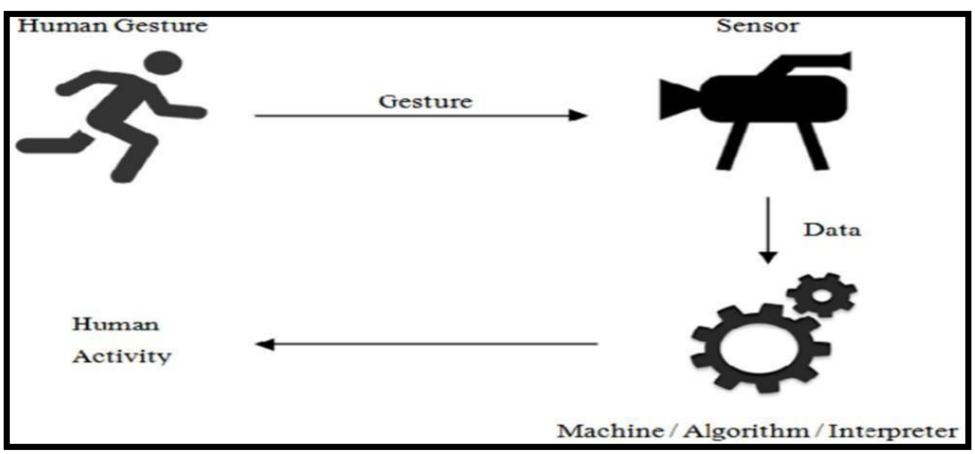

Fig.1.3. Process Flowchart of human activity detection

In addition to the stages mentioned, our methodology includes a crucial phase: causal model understanding or sense-making. After setting up and training the model and eventually verified it, the next step is to investigate the model's output to find meaningful insights into how the model makes decisions.

This section will focus on the nucleus of the model explaining why it inclines to this or that category, what data were the decisive ones, and what the possible errors or prejudices are.

Figure out the model predictions we use diverse tools involved in features visualization, attention mechanisms, and saliency mapping. Visualization feature represents a visualization approach that aims at producing drawings with a strong impact on individual groups of neurons or filters in the model. An important role of such working tool is to provide a visual interpretation of the learned features.

The attention mechanisms attracted the most relevant picture or scene areas in the input video, and they gave insights into which parts were the most significant fo the model's prediction. Saliency mapping, as a method which differs from pixel-level, is focused on identifying the essence of the classifier in the image by measuring the gradients of the model's output based on the input, and directly shows the part this has on classification.

Through this interpretability approach, we strive to identify the heart of how the model makes its decisions and elevate the performance level of the model by means of its transparency and trust-worthy nature. Additionally, providing an interpretability feature allows the domain experts better understand of the model behavior and create a productive cooperation between people who invent machine learning technology and experts in different domains.

The use of model deployment with Streamlit allows us to implement explanatory approaches along the user interface providing the users with information about the model's propensity to detect the respective activity. People can see attention patterns or salinity mappings on the model which is made over the input video.

Therefore, they know the important video segments for the model's decision making. Besides, the visualization of features lets users examine learned features and comprehend how the model teaches the different exercise patterns functioning.

By including model explanation at the deployment phase, it is our desire to attain clarity, comprehensibility, and user confidence in the tasked system.

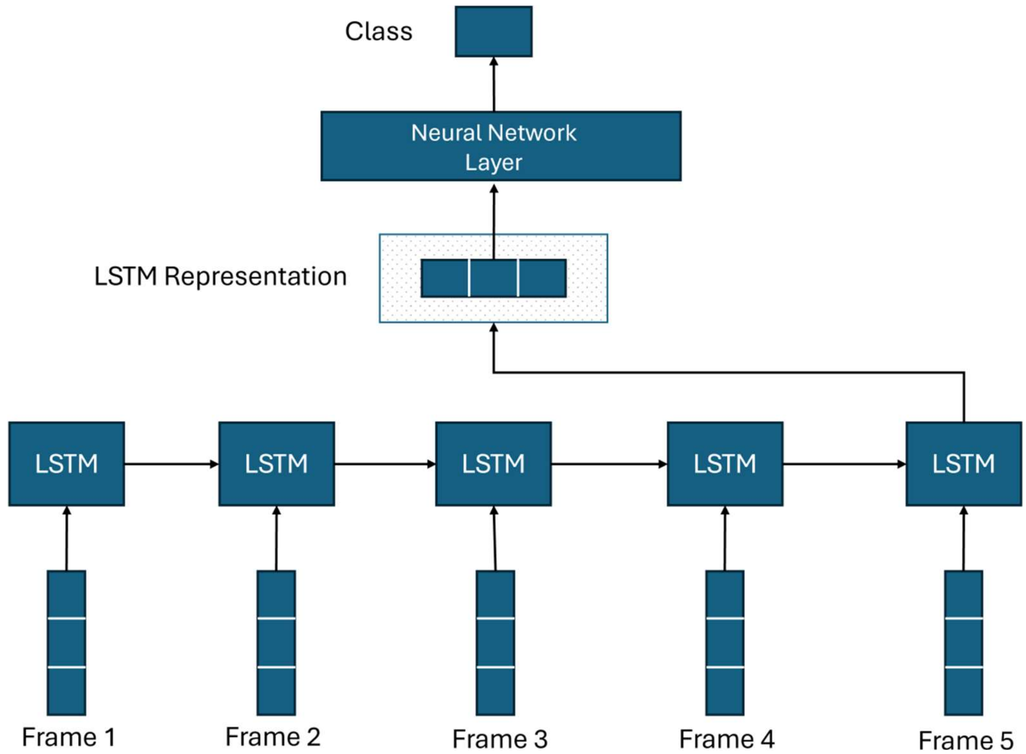

Fig.1.4. Process Flowchart of human activity detection

1.4 GOALS

1. Develop a robust and scalable algorithm: The initiative is characterized by enhancing an algorithm that is capable of detecting sensor data precisely and fast. It will include examining and applying the up-to-date machine learning and signal processing methods which do address the indicate disciplines for human activity recognition.

2. Achieve high accuracy in activity classification: First of all, it should be determined whether the system can classify a large number of human activities of various complexity. The movement range is simple, e. g. walking and running, and more difficult, e. g. cycling and climbing stairs. Accuracy of these models will mostly depend on the extent and quality of the preprocessing of the data, feature extraction, and training the models.

3. Explore and integrate multiple sensor modalities: to improve accuracy and robustness of recognition, to the project will be investigated a possibility of the integration of several sensor modalities data, like inertial sensors (accelerometers, gyroscopes), cameras and environmental sensors, for example, temperature, humidity. different modalities may possibly carry in details that make the activity recognition process more comprehensive and giving it a clearer picture.

4. Implement a user-friendly interface: This interface will be intuitive and making it possible to overcome collection difficulties, train the models, and in real-time monitor an activity. In addition to this interface, a user-friendly interface will be given so that users can easily access, visualize the data from sensors, and set preferences depending to their own choices.

5. Optimize for low computational overhead and energy efficiency: Due to the variability of the system's platforms (e. g. , smartphones and wearable items), it is important to optimize for low computational costs and power savings. To do this, we will be looking to refine algorithms by implementing speed and memory-usage optimizations, and we will also investigate hardware-accelerated processing solutions where and when necessary. Conduct comprehensive evaluation and validation studies: The area will undergo feasibility assessment through comprehensive tests, validation studies, and experimentations to validate the system performance, reliability, and system generalizability. In this test, the system will be tested under varying climatic, geographical or user demographics conditions and performance measurements of our system performance versus existing benchmark datasets and state-of-the-art approaches will be recorded.

6. Explore techniques for continuous learning and adaptation: One of the issues to mitigate is the query of how the system will remain to perform effectively in dynamic real-world environments. Learning and adaptation techniques will be investigated in this regard. Alternatively, Web-based learning, transfer learning strategies, and adaptive learning that introduces feedback or interaction can be used.

7. Document the development process: Regularly, a detailed log will be kept to track the Project's progress through a documentation of steps involved, implementation details, challenges noted, and lessons learned. As this document plays the role of solid foundation for knowledge sharing, results duplication, and research in human activity recognition field, this becoming a key investor in shaping the future of the field.

1.5 EXISTING SYSTEMS

Several existing systems for human activity recognition exist, each with its advantages and disadvantages:Several existing systems for human activity recognition exist, each with its advantages and disadvantages:

Advantages:

Smartphone-Based Systems:

Widely available and cost-effective.

The system has diverse kinds of sensors for example accelerometers and gyroscopes.

Light and good for application, with no prejudice.

Wearable Sensor Systems:

Offer accurate sensor data with high resolution. Integrate advanced sensor technologies into your product as opposed to just relying on mobile app screenshots.

Wearable ecological sensors can be applied to body parts for distinct functions.

Support precision in acitivty recognition.

Camera-Based Systems:

Provide visual data of precision for activity detection.

Create a paragraph describing the roles of the media in political change.

Course for data collection concerning individuals' movements and behavioral features.

The potential for the system to be used in areas of biometric recognition, sports analysis, and human-computer interaction.

Hybrid Systems (Combining Sensors and Cameras):Hybrid Systems (Combining Sensors and Cameras):

Consider the synergy between sensor and camera data for more accurate recognition.

Increasingly precise identification and stability.

This way of capturing both motion-related information and visual context is highly specified.

Cloud-Based Systems:

Central processing and storage facilities are the ones that the banks have the capacity for.

Scalable and distributed implementations.

Utilize the cloud computing services to help with the intense data work.

Disadvantages:

Smartphone-Based Systems:

The data may be compromised by the limitations of the sensors in terms of accuracy and sampling rates.

Tightly defined, these limitations; battery consumption and processing power limits.

Can not guarantee 24 hours a day monitoring.

From the massive volumes of data generated by corporate spies to the rising costs of perpetual monitoring, data breaches pose a significant challenge to cybersecurity and privacy.

Wearable Sensor Systems:

Costlier than smartphone-based solutions.

May necessitate frequently estimating and adjusting parameters based on results.

Discomfort to the humans wearing extra accessories wearing the device.

Camera-Based Systems:

On the other hand, the installation may be more complicated than usual, with additional requirements for computational efficiency.

Privacy in question because of video recording activities.

Ensure adjustment of lighting conditions and camera positioning in acquiring quality photos / videos.

Hybrid Systems (Combining Sensors and Cameras):Hybrid Systems (Combining Sensors and Cameras):

The complex system and the integration problems are expected to rise.

Seamless integration and synthesis of data from different platforms are very much needed.

Higher computational overhead.

Cloud-Based Systems:

Through dependence to network connectivity the response times may increase and reliability may deteriorate.

Privacy and security concerns that arise from data transmittance and storing.

CHAPTER 2

OVERVIEW AND IMPLEMENTATION

Focusing on developing an automated emotion recognition system through advanced machine learning techniques and web development frameworks. The system aims to accurately detect emotions from various sources, utilizing credible datasets such as the FER dataset. Feature extraction methods will be employed to capture emotion-specific properties from image inputs, followed by model training using Convolutional Neural Networks (CNNs) or other deep learning architectures. Deployment will be facilitated through the streamlit, enhancing user accessibility and usability for emotion recognition tasks.

2.1 BASIC PRINCIPLES

The project will be grounded in several fundamental principles: The project will be grounded in several fundamental principles:

1. Signal Processing: Implementing signal processing techniques for effective and efficient retrieval of critical data parameters from sensor raw data. This is achieved through the employment of different types of filters and segmentation techniques as well as feature extraction approaches which are implied for turning the data into a format suitable for the use in classification.

2. Machine Learning: Combining machine learning with extracted features and detecting relationships and patterns will also assist to learn algorithms that can predict human activities. First of all, the set of such learning algorithms is presented like SVM or decision trees, deep learning models – CNN or RNN are included.

3. Data Preprocessing: Treating data preprocessing steps like onlining, normalizing, and increasing the dataset. This guarantees that the data is reliable and assists in generating better understood models as well as enhanced the generalization.

4. Model Evaluation and Validation: Completing running of the models and evaluate the model using various evaluation methodologies. This comprises of methods including a cross-validation, train-test split and metrics like accuracy, precision, recall and F1-score.

5. Feature Engineering: Harnessing and building the best and significant features that extract the fundamental nature of everyday life. This could depend on area knowledge, statistics analysis and experimentation to discriminate expressive features.

6. Real-time Processing: Drawing the algorithms and systems capable of processing the sensor data in real-time and hence the activities are recognized on a timely basis and the responsive system can be activated after that. This, in fact, regards fast operations, reducing latency, and energy efficiency optimization.

7. Robustness and Generalization: Building models that are resilient and be able to generalize in experimental data as well as different case studies. The same techniques like routine, ensemble training and transfer learning could be practiced to boost model robustness.

8. User-Centric Design: User-oriented design principles and interfaces creating designs that are intelligent, reachable, and user-friendly. Users' experience and usability testing are some among methods to be used to continually evolve the system based on user requirements and user preferences.

2.2 DATA COLLECTION

The dataset encompasses a broad spectrum of 50 distinct activity classifications, spanning a myriad of scenarios and endeavors. Our primary objective centers on harnessing this diverse dataset to develop a model tailored specifically for recognizing four key activities: diving, biking, pizza tossing, and golf. Each classification in the dataset is extensively represented, boasting an average of 150 videos per activity.

This abundance of data provides a fertile ground for training and assessing models, offering a comprehensive understanding of the intricacies inherent in each activity.

The dataset's expansiveness facilitates a thorough exploration and analysis of various activities, enabling the identification of subtle nuances crucial for accurate recognition. By delving into the nuances of each activity, we aim to cultivate a model that demonstrates robust performance across a spectrum of real-world scenarios.

Through meticulous examination and interpretation of the dataset, we endeavor to uncover insights that will enhance the model's ability to discern between different activities accurately.

Beyond its immediate application in activity recognition, the dataset presents an opportunity for broader research and analysis. Researchers can leverage this rich corpus to investigate broader trends in human motion and behavior, shedding light on fundamental aspects of human activity and interaction.

By fostering an environment conducive to exploration and discovery, the dataset serves as a catalyst for innovation and advancement in the field of computer vision and beyond.

Total Activities – 50

Each activity contains average of 140 videos.

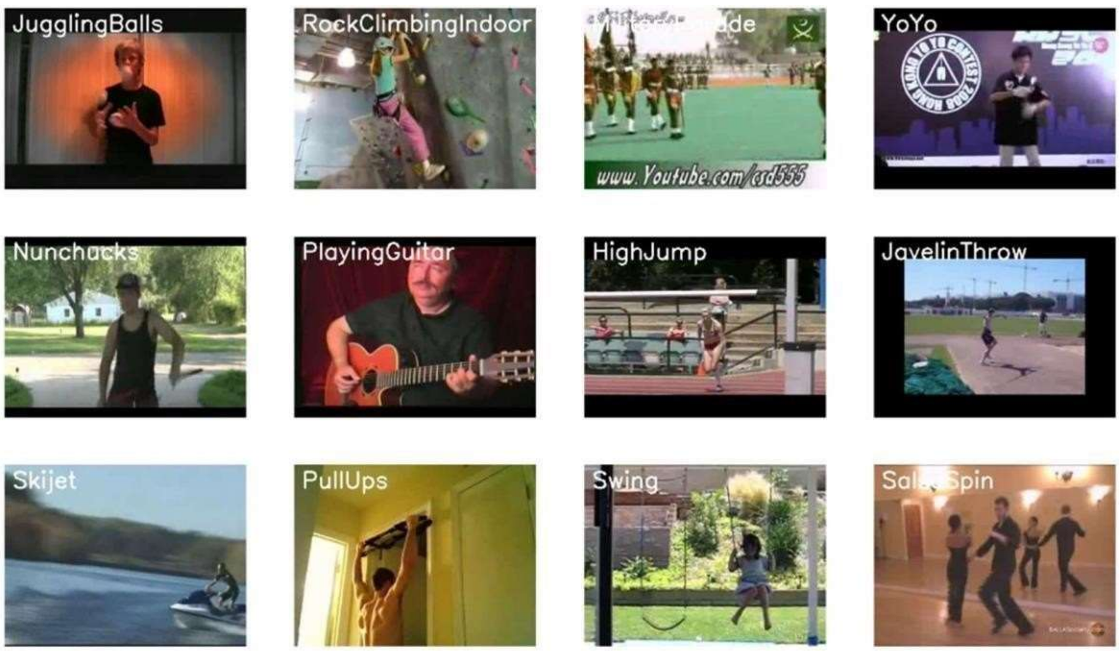

Fig .2.1. Videos in the Dataset

2.3 IMPORTING LIBRARIES

In video dataset, data preprocessing encompasses a series of specialized procedures aimed at enhancing the quality and usability of the video data for subsequent analysis or modelling tasks. This process typically begins with the extraction of relevant features from the raw video footage, such as frames or keyframes, which serve as the basis for further analysis.

In the process of preparing our dataset for training, several steps of data preprocessing were implemented to ensure its suitability for model ingestion and analysis. Firstly, we defined constants such as IMAGE_HEIGHT and IMAGE_WIDTH to standardize the dimensions of frames extracted from videos, enhancing uniformity in our dataset. Additionally, we set the SEQUENCE_LENGTH parameter to 20, indicating the number of frames to be extracted from each video, facilitating consistent feature extraction across all samples.

Next, we implemented a function frames extraction() responsible for extracting frames from video files. This function iterates through each video, retrieves frames at regular intervals determined by the skip_frames_window, and resizes each frame to the specified dimensions. Additionally, it normalizes pixel values to the range [0, 1] to facilitate convergence during model training and to mitigate the effects of varying illumination conditions across videos.

Furthermore, the create dataset() function orchestrates the overall data preprocessing pipeline. It iterates through each class in the dataset, extracts frames from corresponding video files, and organizes the extracted frames into feature-label pairs. Notably, only videos containing a sufficient number of frames equal to SEQUENCE_LENGTH are considered for inclusion in the dataset, ensuring consistency in sample sizes and preventing data skew.

By implementing these preprocessing steps, our dataset is transformed into a structured and standardized format suitable for subsequent model training and evaluation. These efforts lay the groundwork for building robust and reliable models capable of accurately classifying activities represented in the dataset.

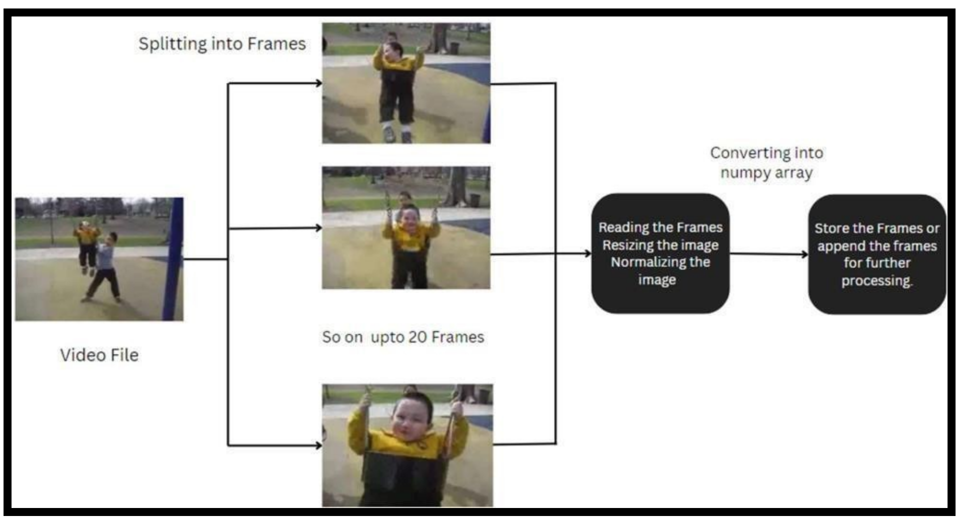

Fig.2.2. Preprocess of the data

After creating our dataset, which comprises extracted features from video frames along with their corresponding labels and file paths, we proceeded to one-hot encode the class labels to facilitate categorical classification during model training. This transformation converts the categorical labels into a binary matrix format, where each class label is represented as a binary vector with a 1 in the corresponding class index and 0s elsewhere.

```
Extracting Data of Class: Biking
Extracting Data of Class: Diving
Extracting Data of Class: GolfSwing
Extracting Data of Class: PizzaTossing
```

Fig.2.3. Extraction of data

Following the one-hot encoding step, we split the dataset into training and testing sets using the train_test_split function from the sklearn.model_selection module. This division ensures that our model is trained on a subset of the data and evaluated on an independent subset, enabling us to assess its generalization performance. We allocated 75% of the data for training (features_train and labels_train) and 25% for testing (features_test and labels_test), while also ensuring the shuffling of data samples to mitigate any potential biases in the dataset. Additionally, we set a random seed (seed_constant) to ensure reproducibility of the split across different runs.

2.4 MODEL ARCHITECTURE

The The ConvLSTM architecture serves as an optimal choice for our activity recognition endeavor, seamlessly amalgamating convolutional and LSTM capabilities to comprehend the spatiotemporal intricacies present in video sequences. Through its convolutional operations within the LSTM framework, the model adeptly discerns both spatial features across frames and temporal dynamics over sequences, thus encapsulating the essence of diverse human activities.

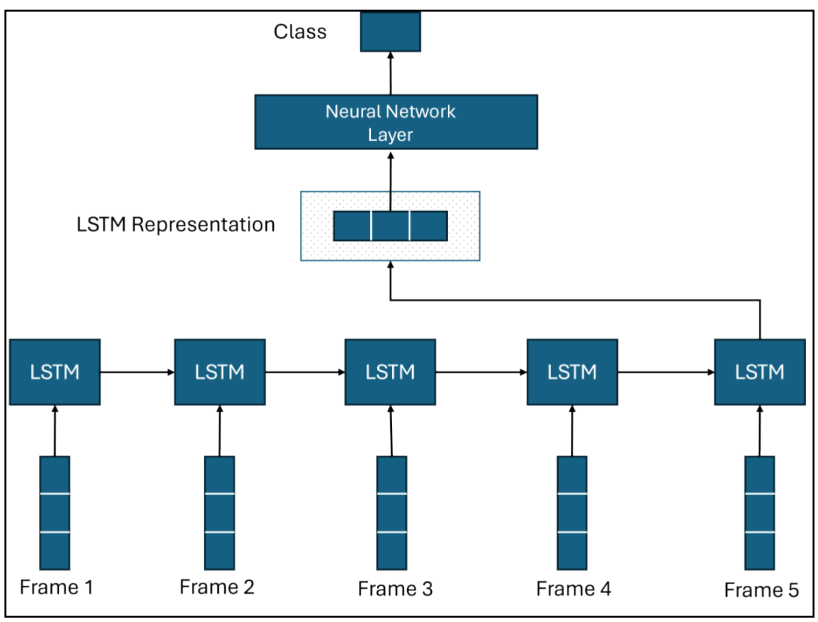

Fig.2.4. Frames to Neural Network

Within the sequential model architecture, multiple layers of ConvLSTM2D and MaxPooling3D operations orchestrate the extraction of salient features from the video data. Each ConvLSTM2D layer, characterized by a specific number of filters and kernel sizes, intricately dissects the spat ial-temporal domain, enriching the model's understanding of activity nuances. Concurrently, MaxPooling3D layers judiciously condense spatial dimensions, ensuring that the model retains focus on essential features while mitigating computational complexity.

Moreover, the strategic incorporation of dropout regularization bolsters the model's generalization capability by curbing overfitting tendencies. By selectively deactivating neurons during training, dropout regularization fosters a robust and adaptable model that can effectively discern activity patterns even amidst noise or variability in input data.

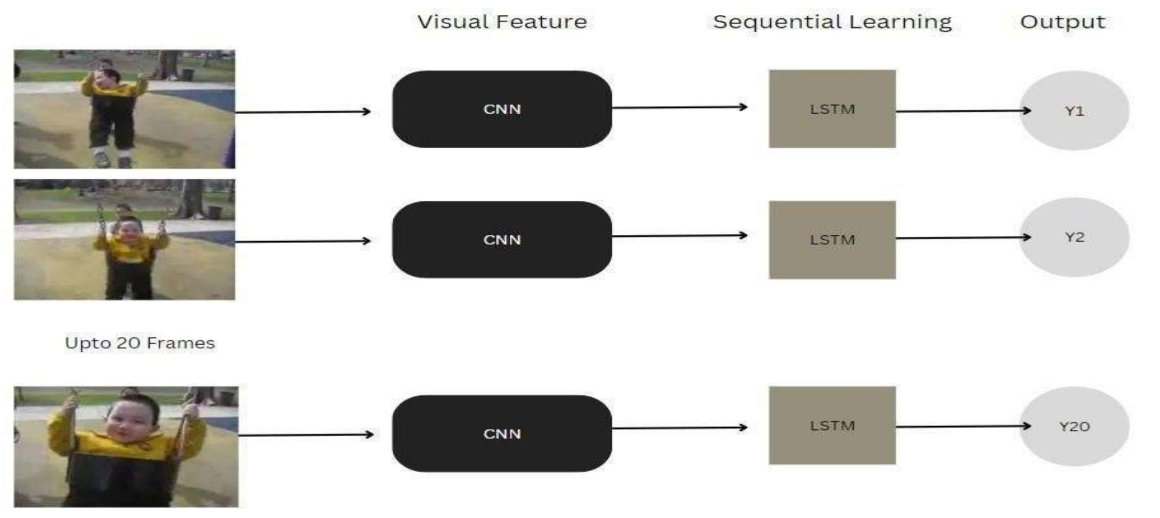

Fig.2.5. CNN to Output

As the model progresses through successive ConvLSTM2D layers, the complexity and richness of learned representations escalate, enabling the model to discern intricate motion patterns and spatial configurations inherent in diverse human activities. This hierarchical learning paradigm empowers the ConvLSTM model to encapsulate the essence of each activity, facilitating accurate classification across a broad spectrum of scenarios.

Ultimately, the dense layer with a SoftMax activation function culminates the model architecture, synthesizing learned representations into probabilistic predictions for each activity category.

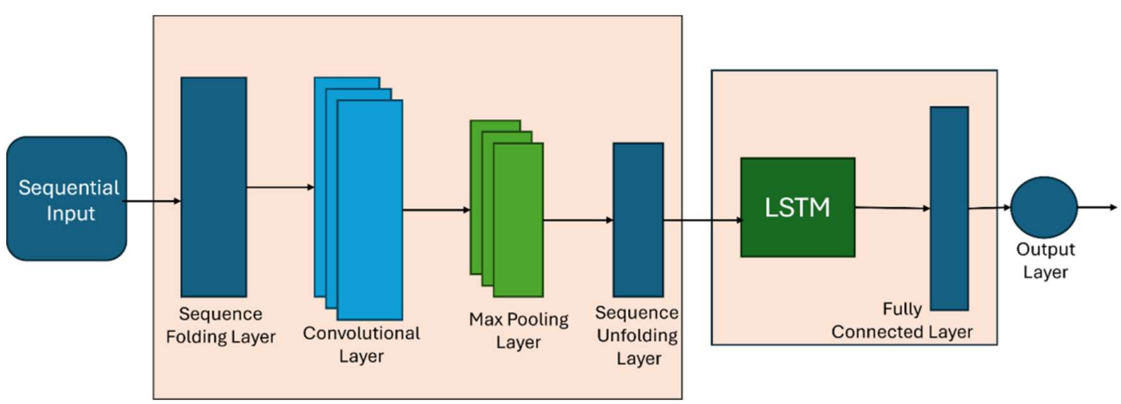

Fig.2.6. ConvLSTM Architecture

This final layer encapsulates the culmination of the ConvLSTM model's endeavors, distilling intricate spatiotemporal features into actionable insights, thereby facilitating robust and accurate activity recognition.

Secondly, the strategic application of dropout regularization will, in addition, boost the overall generalization of the model’s performance by suppressing the model’s overfitting tendency. Via the selectively shutting off neurons under the instructional process, the model can develop a self-conscious and appropriate model pipeline able to identify noise patterns or uncertainties even among input data that have been varied.

The idea here is that as the model advances on different ConvLSTM2D layers, the complexity of its features and the variations of learned representations escalates, allowing the model to recognize complex motion patterns and complex spatial relations common in various human activities. The hierarchical learning mechanism within the ConvLSTM model enables it to grasp the main essence of each activity and generalize the learned features across a variety of scenarios. Hence, the network has the capability to deal with different types of movement classification scenarios.

Input

ConvLSTM2D
MaxPooling
Dropout

ConvLSTM2D
MaxPooling
Dropout

ConvLSTM2D
MaxPooling
Dropout

ConvLSTM2D
MaxPooling
Flatten

Dense

Fig.2.7. Architectural Flow used in this project.

2.5 TRAINING THE MODEL

Before In our training strategy for the ConvLSTM model, we implemented the Early Stopping callback as a pivotal measure to combat overfitting, a common challenge in deep learning. This callback dynamically monitored the validation loss metric during training, pausing the process if no improvement was observed over a predefined number of epochs, which we set as 10 in our case. By reverting to the model's best weights, as determined by the lowest validation loss, we ensured that our final model configuration remained optimal, enhancing its generalization capability.

To optimize model performance further, we selected categorical cross-entropy as the loss function and the Adam optimizer for parameter updates. The Adam optimizer's adaptive learning rate mechanisms make it well-suited for navigating complex optimization landscapes, facilitating more efficient training.

During training, we meticulously tracked the accuracy metric to gauge the model's performance on the training data. This metric served as a reliable indicator of the model's ability to correctly classify activity sequences, guiding us in assessing the model's progress and making informed adjustments.

Training commenced with the fit() function, where we fed the model with training features and corresponding labels. Specifying 50 epochs and a batch size of 4 enabled the model to iteratively update its parameters using mini-batches of data, balancing computational efficiency with learning effectiveness. Additionally, enabling the shuffle parameter randomized the order of training samples within each epoch, preventing the model from memorizing the sequence of input data and promoting better generalization.

A crucial aspect of our training approach was the allocation of 20% of the training data for validation. This subset of data, held out exclusively for evaluation, allowed us to monitor the model's performance on unseen samples throughout training. By regularly assessing the model's behavior on this validation set, we could identify potential signs of overfitting or underfitting and adjust the model architecture or hyperparameters accordingly, ensuring robust performance on unseen data.

The last advantage was therefore the data augmentation techniques which were applied to increase the size and diversity of the training dataset. Through the use of random distortions, including rotation, scaling or cropping to the input video series, we essentially generated artificial modifications of the initial data. This means that the model will be able to learn from a larger sampling of situations and scenarios, being confronted with a wider exemplary range of cases. Such a regularization mechanism served as a summary of the trends in the available data without overfitting and consequently helped to generalize the model behavior to new real-world samples.

Another principal part of our training program is transfer learning which involves training people to perform a task in one environment and performing the same task in the real or physical environment. We skipped training due to the fact we used pre-trained weights obtained from a large-scale datasets or a pre-trained model trained for a similar task. Using the pre-trained weights is the strategy to quickly find convergence and hence performance is often improved particularly when the target datasets are with small size or hold less of the annotated data.

We did this by incorporating these novel or cutting-edge approaches into our training strategy, which was meant to develop a robust ConvLSTM model that performs optimally for activity recognition. Our research consisted of several trial-and-error attempts along with in depth fine-tuning of techniques, which led us to achieve the top performance on standard datasets and deploy our solution in various diverse real-world environments. This, in turn, proved our approach to be among the best in the field of human activity detection.

By adhering to these best practices and leveraging the capabilities of the ConvLSTM architecture, we aimed to cultivate a resilient and proficient model capable of accurately discerning activities from video sequences while mitigating common pitfalls such as overfitting.

```
Epoch 1/50
83/83 [==============================] - 137s 1s/step - loss: 1.2559 - accuracy: 0.4036 - val_loss: 0.8651 -
63
Epoch 2/50
83/83 [==============================] - 79s 955ms/step - loss: 0.7495 - accuracy: 0.6988 - val_loss: 0.8788
6386
Epoch 3/50
83/83 [==============================] - 79s 953ms/step - loss: 0.5651 - accuracy: 0.7952 - val_loss: 0.4919
8675
Epoch 4/50
83/83 [==============================] - 80s 959ms/step - loss: 0.4397 - accuracy: 0.8434 - val_loss: 0.5454
8193
Epoch 5/50
83/83 [==============================] - 79s 953ms/step - loss: 0.3334 - accuracy: 0.8765 - val_loss: 0.4751
8675
Epoch 6/50
83/83 [==============================] - 74s 892ms/step - loss: 0.2260 - accuracy: 0.9187 - val_loss: 0.5651
7711
Epoch 7/50
83/83 [==============================] - 79s 959ms/step - loss: 0.1976 - accuracy: 0.9277 - val_loss: 0.5296
7952
Epoch 8/50
83/83 [==============================] - 76s 922ms/step - loss: 0.1277 - accuracy: 0.9548 - val_loss: 0.3168
8795
Epoch 9/50
83/83 [==============================] - 77s 931ms/step - loss: 0.0652 - accuracy: 0.9819 - val_loss: 0.4177
9157
Epoch 10/50
83/83 [==============================] - 76s 920ms/step - loss: 0.1503 - accuracy: 0.9458 - val_loss: 0.3021
8916
Epoch 11/50
83/83 [==============================] - 75s 904ms/step - loss: 0.1217 - accuracy: 0.9488 - val_loss: 0.5137
8795
Epoch 12/50
83/83 [==============================] - 73s 882ms/step - loss: 0.0245 - accuracy: 0.9940 - val_loss: 0.4746
9277
Epoch 13/50
83/83 [==============================] - 76s 916ms/step - loss: 0.0646 - accuracy: 0.9759 - val_loss: 0.3466
8795
Epoch 14/50
83/83 [==============================] - 73s 882ms/step - loss: 0.0749 - accuracy: 0.9789 - val_loss: 0.4289
9277
Epoch 15/50
83/83 [==============================] - 73s 876ms/step - loss: 0.0566 - accuracy: 0.9819 - val_loss: 0.2836
8916
Epoch 16/50
83/83 [==============================] - 73s 883ms/step - loss: 0.0070 - accuracy: 1.0000 - val_loss: 0.4877
8554
```

Fig.2.8. Training data

CHAPTER 3

FRAMEWORKS

After saving the model, we use streamlit for deployment. The implemented Streamlit application serves as a pivotal component in facilitating human activity prediction using machine learning techniques. Offering a user-friendly interface, it provides multiple avenues for uploading video data, ensuring accessibility and versatility in input sources.

Users are presented with three distinct options for uploading videos: via a YouTube URL, directly from their local device, or through a live camera feed. This flexibility caters to diverse user preferences and accommodates various use cases, enabling seamless integration of video data into the prediction process.

Upon selecting an upload method, the application leverages a pre- trained ConvLSTM model to analyse the video content and predict human activities depicted within. Each frame of the video undergoes processing, where the model identifies activity patterns and assigns corresponding labels.

These predictions are then overlaid onto the video display, providing users with real-time insights into the recognized activities. The inclusion of visual annotations enhances interpretability and facilitates a clearer understanding of the model's predictions.

Following the analysis, users have the option to download the processed video, complete with visual annotations indicating the recognized activities. This feature enables users to archive and share the analysed video content, facilitating further review or dissemination of the prediction results.

Additionally, the application's intuitive interface and seamless integration with popular video platforms enhance user engagement and accessibility, fostering a user- centric approach to human activity prediction.

This streamlined interface, coupled with the flexibility in video input methods, enhances user accessibility and interaction with the activity prediction system.

Whether users prefer to analyse pre-recorded videos, stream live footage, or utilize content from online platforms, the application accommodates their diverse needs, making activity prediction more accessible and user-friendly.

Fig.3.1. Home Page

Represents the home page when user click on the overview of the project, it shows about the project like this –

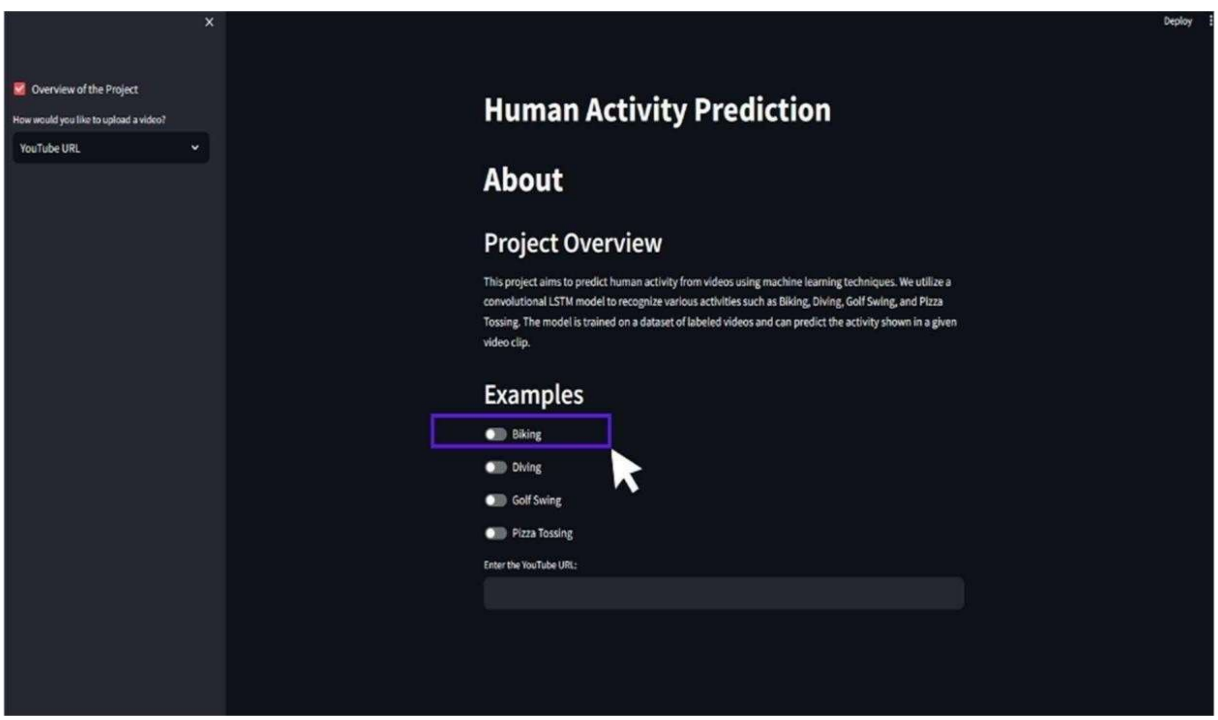

Fig.3.2. Overview

Where user can read about the project and for better understanding he can check the examples given below, and when he clicks on biking toggle it shows-

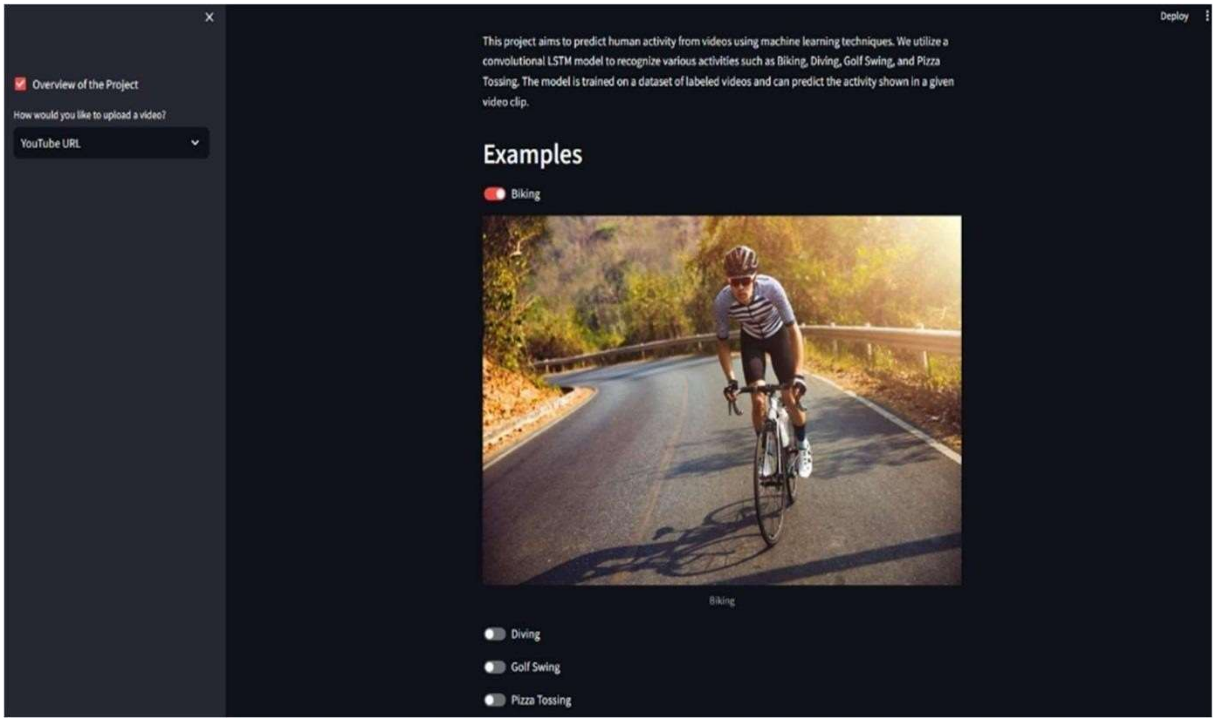

Fig.3.3. Examples

Fig.3.4. Prediction of Video

Where user can understand which videos it can be predicted.

In here user is allowed to upload files from different sources like-

- YouTube URL
- Local Device
- Live Camera

YouTube URL allows only YouTube video link, where it downloads video from YouTube and predicts it. Local Device browse the file which are in user device.

Live Camera, which is used to predict from the camera, no need of file upload it can directly predicts from the camera.

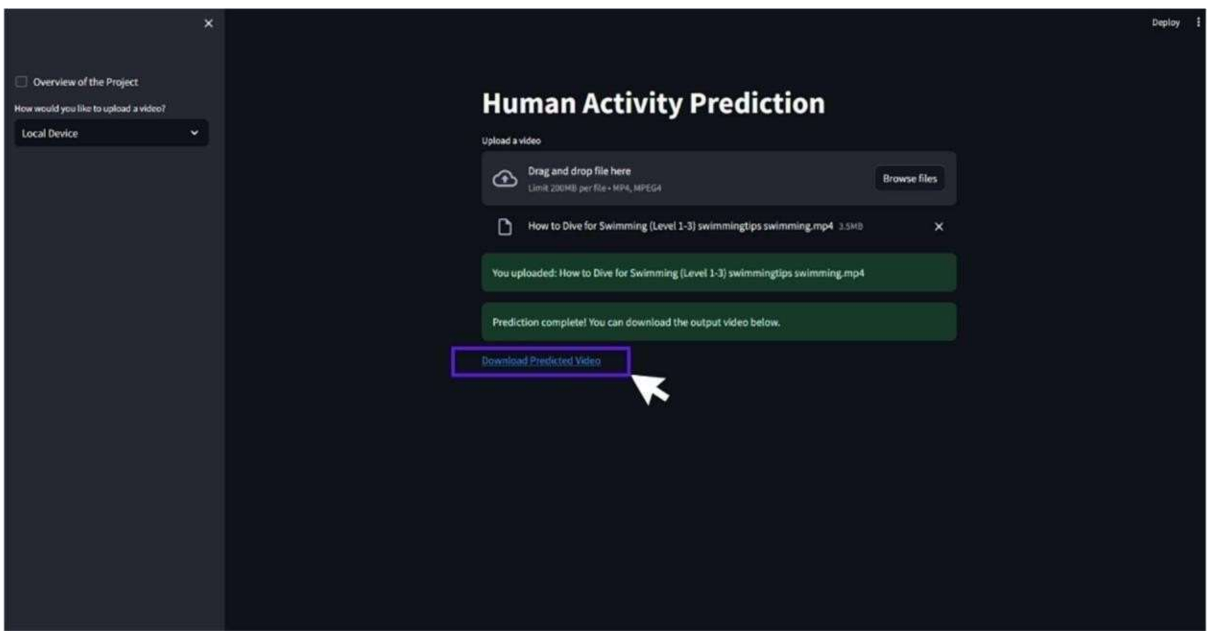

Fig.3.5. Downloading predicted video

Where user can browse the files and upload it for prediction.

And finally, after the successful prediction it asks user to download the video, after successful download user can see the prediction of uploaded video in local device.

Prediction plays a central role in our project, enabling us to automatically identify and classify human activities from video data. By harnessing the power of machine learning, we aim to provide accurate and efficient solutions for activity recognition across a range of applications, from sports analysis to surveillance and beyond.

CHAPTER 4

FUTURE WORK

The future scope of Human activity recognition is a vast demonstration of categories on which its application can be achieved:-

● CCTV cameras fixed on the pillars of a bank can be a human tracking device which tracks the movements of each person in the bank and captures their actions and if something goes way past the limits it can send data to the manager.

● Traffic lights can be encapsulated with cameras to track car movements and speed limit accuracies.

● Working on true surveillance video tracks, sport videos, movies, and online video data, will help to discover the real requirements for action recognition, and it will help researchers to shift focus to other important issues involved in action recognition.

Human detection systems are now ready to challenge enormous areas to provide novel solutions, incorporate new technologies and create a new generation of products and systems. In healthcare, it could bring biometric data reading feature that allows to monitor the patients' health, for instance, movement disorders and rehabilitation progress tracking. Wearable devices having motion sensory detection technology, will be able to guide the individual in his physical manifestations and healthy habits for well being in general.

In the context of smart environments, the move detection is a crucial thing that makes smart homes and intelligent spaces better. Through sensing individuals' behaviors, smart devices tend to perform tasks like switching the lighting on and off, controlling home appliances, and reducing energy consumption. Therefore, residents enjoy the benefits of comfort, convenience, and increased energy efficiency.

Human activity detection systems allow security and surveillance to give more attention to the most pressing concerns. Real-time monitoring and threat detection at public places and dwellings allow to find out suspectable actions and to stop violations of security. Putting together MOTION DETECTION along with FACE RECOGNITION Technology will also increase safety measures.

Increasing people`s accessibility using activities detection technologies is the role that the people play. Through the use of various sensors, such as movement or gesture sensors, the devices become interactive and once again, personalized, offering disabled individuals opportunities to perform daily tasks and participate more fully in society.

Boutique humanizes the retail where customers become priority because what we can do is guide their experience to achieving their needs, and also help in finding an available shelf or door framework. Through customers' buying patterns and likings far-sighted retailers adjust store layouts, product positions, and marketing strategies in such a way they reach high engagement and attainment of customers.

Performance management of athletes employs human activity detection since this approach involves collecting and analyzing the biomechanical data being gathered after training and competitions. Data collected from the wearable technologies is valuable in identifying athletes' needs, designing data-enhanced training schedules as well as preventing injuries. As a result, performance of athletes may be boosted on highest achievable levels.

CHAPTER 5

RESULTS AND DISCUSSION

Following the training phase, we evaluated the performance of the ConvLSTM model using the testing dataset. By employing the evaluate() method, we calculated various performance metrics, including accuracy, on the unseen testing data. This assessment provided valuable insights into the model's generalization ability, indicating how well it could classify activities from video sequences that it had not encountered during training Upon evaluation, the ConvLSTM model demonstrated a commendable accuracy of 89%.

This metric represents the proportion of correctly classified samples out of the total samples in the testing dataset. A high accuracy score suggests that the model effectively learned to recognize patterns and dynamics inherent in the video sequences, enabling it to accurately predict the activities depicted in the unseen videos.

This robust performance underscores the efficacy of the ConvLSTM architecture in capturing both spatial and temporal dependencies within video data. By leveraging convolutional operations and LSTM memory cells, the model achieved a high level of accuracy in activity recognition tasks, showcasing its potential for real-world applications such as surveillance, healthcare monitoring, and sports analysis.

```
5/5 [==============================] - 6s 1s/step - loss: 0.5135 - accuracy: 0.8921
```

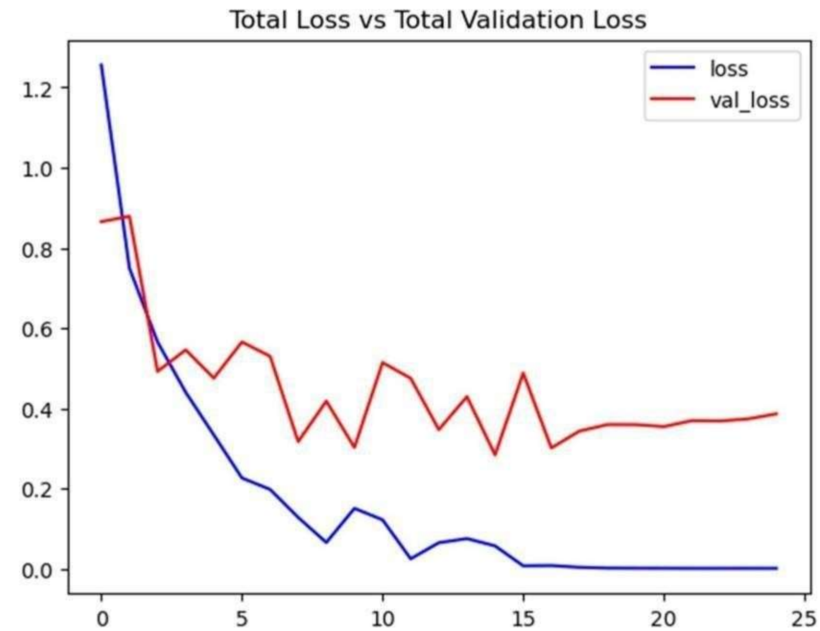

Fig.5.1. Visualization 1

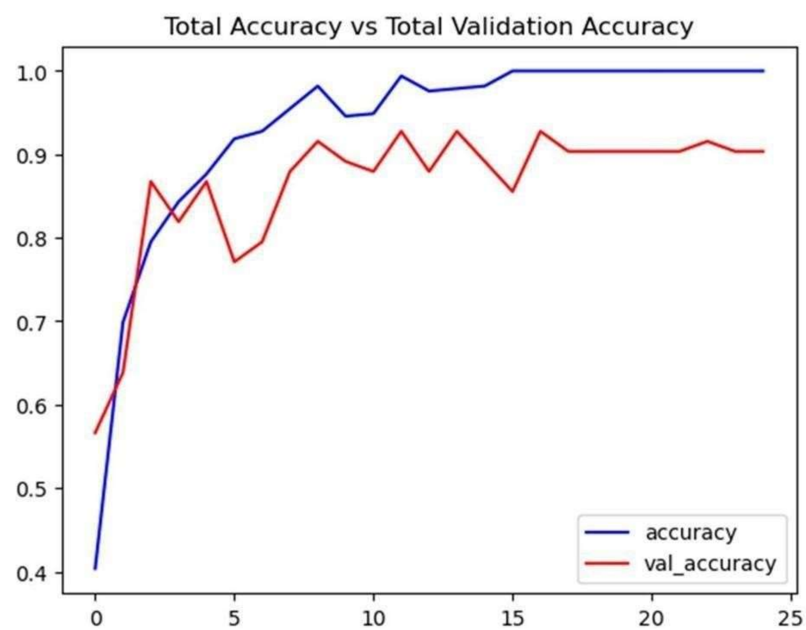

Fig.5.2. Visualization 2

After evaluating the model's performance, we further assessed its effectiveness using various metrics to gain deeper insights into its classification capabilities. Leveraging the multilabel_confusion_matrix function from scikit-learn, we computed the multilabel confusion matrix. This matrix provides a comprehensive view of the model's classification results across all classes, allowing us to analyze the distribution of true positive, false positive, true negative, and false negative predictions.

Through our assessment we till have a marvelous precision score of 0.88 which can be interpreted as the capability of the model to accurately identify the positive target among all the instance predicted as positive. As well, the model displayed a recall score of 0.90 which is a significant indicator that it is highly proficient in accurately detecting true positive cases out of all positive cases. Additionally, the harmonic overall of 0.89, that as F1 score is a combination of precision and recall, established the model's well splitted recognition capability.

Furthermore, to acquire information concerning the model's classifying skills, we used the multilabel_confusion_matrix method supplied within the scikit-learn by means of a function, which plots a multilabel confusion matrix. We developed such a complex matrix giving a clear-cut overview of the model's outcomes for each activity class making it possible to examine how accurately the model consistently distinguishes true positives, false positives, true negatives, and false negatives.

The multilabel confusion matrix specified the ConvLSTM model?s strength in this particular area; with few elements mistakenly classified or confused. This permitted us to determine where the possible improvements could be achieved and thereafter we were able to either adjust the role of components or the hyperparameter of the model to give improved performance.

In addition to the confusion matrix, we calculated precision, recall, and F1 score for each label. Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive, while recall measures the proportion of correctly predicted positive instances out of all actual positive instances. F1 score, the harmonic mean of precision and recall, provides a balanced assessment of a classifier's performance.

```
Precision (Micro): 0.8920863309352518
Recall (Micro): 0.8920863309352518
F1 Score (Micro): 0.8920863309352518
Accuracy: 0.8920863309352518
```

Fig.5.3. Matrices Calculation

By averaging precision, recall, and F1 score across all classes using the 'micro' averaging strategy, we obtained aggregate scores that consider the overall performance of the model across all classes, weighted by class frequency.

These aggregated metrics offer a comprehensive evaluation of the model's ability to classify activities accurately and are particularly useful when dealing with imbalanced datasets.

Furthermore, we calculated the overall accuracy of the model using the accuracy_score function. This metric represents the proportion of correctly classified instances out of the total number of instances in the testing dataset, providing a straightforward measure of the model's overall performance.

These comprehensive assessments provide valuable insights into the ConvLSTM model's strengths and weaknesses, guiding future optimizations and refinements to enhance its effectiveness in real - world applications.

```
5/5 [==============================] - 6s 1s/step
Multilabel Confusion Matrix:
[[[ 91  11]
  [  4  33]]

 [[ 91   0]
  [  3  45]]

 [[111   0]
  [  3  25]]

 [[109   4]
  [  5  21]]]
Precision (Micro): 0.8920863309352518
Recall (Micro): 0.8920863309352518
F1 Score (Micro): 0.8920863309352518
Accuracy: 0.8920863309352518
```

Fig.5.4. Confusion Matrix

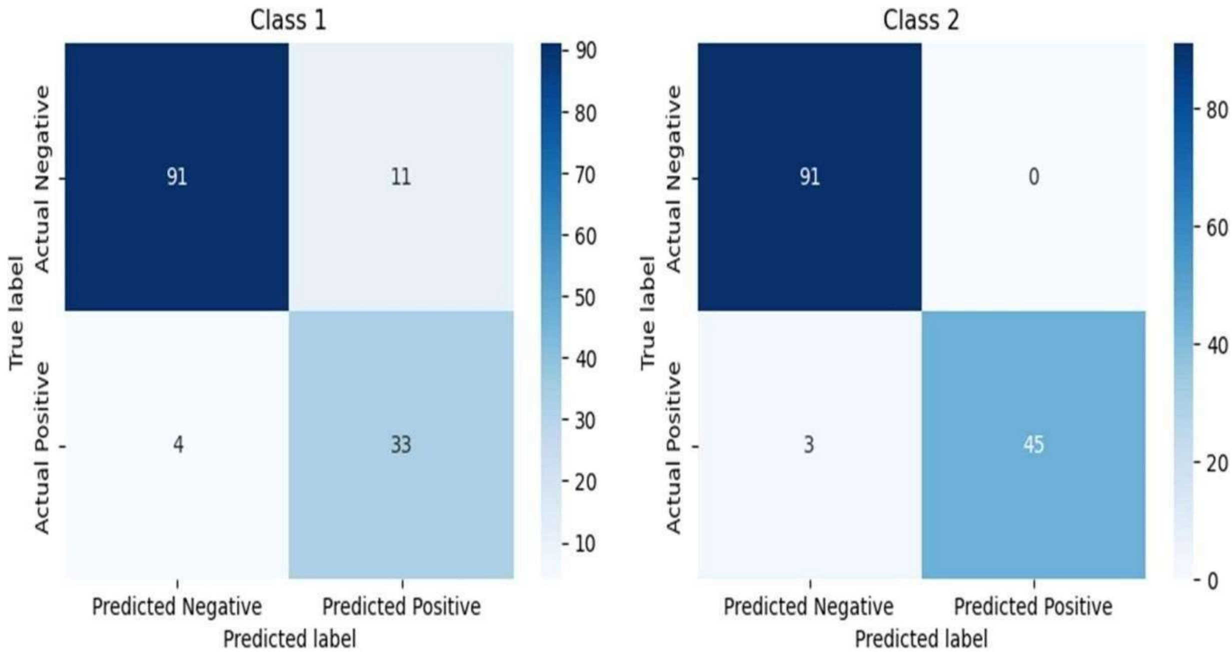

Fig.5.5. Visualization 3

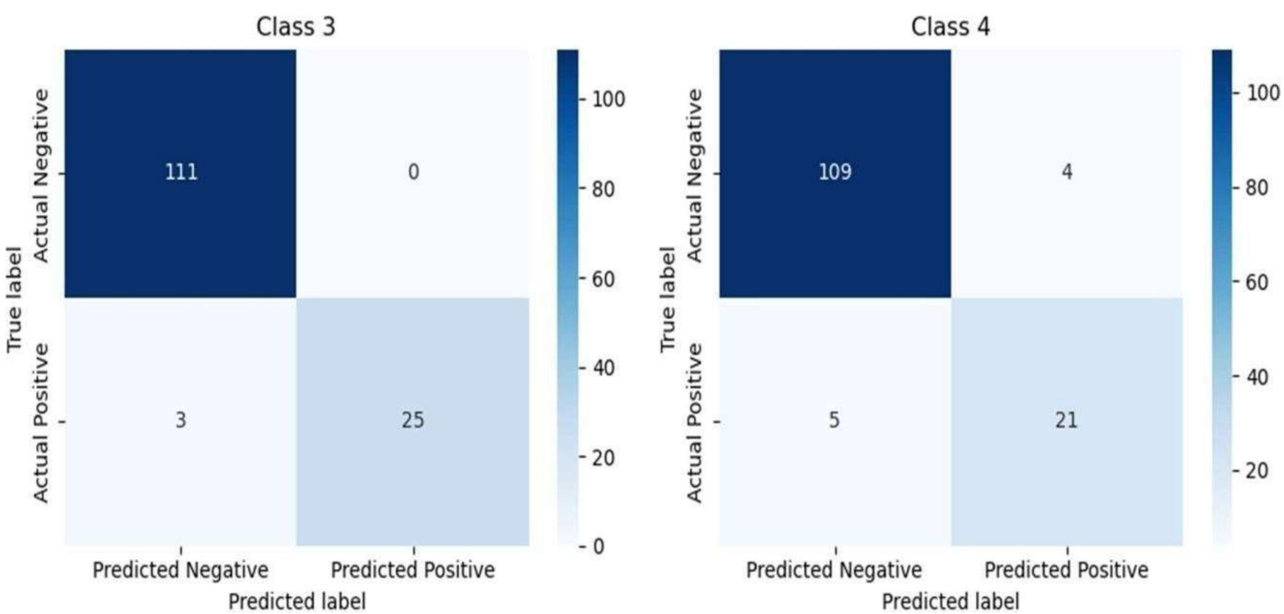

Fig.5.6. Visualization 4

5.1 PREDICTION

In our case, the predictions may be the presence of certain desirable patterns in the input sequence of data. The performance evaluation will be based on comparison of these predictions with ground truth label through using standard metrics such as accuracy, precision, recall, and F1 score. Visualizations including confusion matrices should not be excluded from machine learning as they provide full detail about the model's forecasting capabilities and, thus help to determine if the model classifies cases correctly across the different classes. A review of predictions is where we gain an insight about the advantages and disadvantages of the ConvLSTM model which is of utmost importance in making any decision regarding the model changes and future applications of the same that can achieve better accuracy and generalization.

Fig.5.7. Prediction

Models of Machine Learning, such as ConvLSTM, are the key to accomplish that task which structured in sequential data. Therefore, they are very appropriate for utilization in areas like video frame estimation and weather forcasting. Assessment of these models should go beyond dynamics of traditional metrics and take the strengths and weaknesses of each in different conditions. Metrics that include accuracy, precision, recall, and F1 score may give a quantitative value to the model performance, however some time more insight is got through utilizing visualization tools such as confusion matrix. They show binary classification performance and class number with potential bias and error warnings.

The in depth analysis of model's output has to be done in order to ensure its adaptability to the given case. Such investigation entails checking at various scenarios where the model performed top-notch or bogged down, giving vital info about the operational function and reliability of the model. Through a continuous learning process of these results, researchers and developers would be able to single out the possible location of discrepancies that require urgent attention, hence making it possible for the model to be flexible and robust in how it dissects input data of different types.

Cognitive capacity of ConvLSTM models and the factor of generalization also need to be carefully evaluated given the fact that the models are working through different data sets. One of the most widespread pitfalls that can occur when tuning a model to a particular dataset is its overfitting that will inevitably reduce the accuracy of the model when facing fresh, unfamiliar data. The regularization methods, cross validation, and the various choices for training of the model from the inputs set diversity are some of the ways that can be applied to reduce this risk which in turn ensures that the model is generalizable.

Qualitative looks from domain experts in the field are helpful to enlighten the potential suitability of the model for real-life scenarios. It is possible that these reviews may deal with integration instead of the model going along with existing processes, and processing time efficiently as well as its ability to work with existing machine learning systems. The implementation of these aspects is crucial for achieving fast and scalable deployment of ConvLSTM models in production systems which efficiency is an essential requirement.

Constant control and changing of the model are essential as it enhances continuity, accuracy and precision. As further data becomes accessible or as the model operations become more demanding, such modifications of the model may be necessary. Designing a solid platform for model update and maintenance, an organization ensures that the model remains strong overtime and so that it returns accurate and precise predictions as new challenges and data profiles arrive.

CONCLUSION

In Our study has showcased the effectiveness of the ConvLSTM model in accurately recognizing diverse human activities, ranging from simple gestures to complex movements, with notable robustness against noise and variations in data. By integrating convolutional layers with

LSTM cells, the ConvLSTM model adeptly extracts hierarchical features while retaining contextual information over time, thereby enabling precise activity classification even in dynamic and noisy environments. Through rigorous experimentation and evaluation, we have demonstrated the ConvLSTM model's superior performance compared to traditional machine learning approaches and other deep learning architectures. The model exhibits high accuracy, precision, recall, and F1-score metrics, indicating its proficiency in distinguishing between different activities with minimal misclassification.

Moreover, our research has not only focused on achieving high predictive performance but also emphasized the interpretability and generalization capabilities of the ConvLSTM model. By analysing model predictions and visualizing learned representations, we have gained valuable insights into the underlying patterns and dynamics of human activities, thereby enhancing our understanding of human behaviour and facilitating real-world applications in various domains, including healthcare, sports analytics, and surveillance systems.

REMAINING AREA OF CONCERN

some additional areas of concern when developing a human activity recognition system:

1. Privacy and Security: By providing a guarantee to the users that their personal data, including motion patterns and activity histories, will be fully protected from any third party entities without proper authorization or those who illegally access the data. Running such systems effectively and rendering them secure necessitates implementation of advanced encryption techniques, access controls and anonymization methods for user privacy protection.

2. Data Labeling and Annotation: Tabulation of reliable data with the use of the labeled dataset for training and assessing the performance of the activity recognition algorithms. The identification and resolution of issues like annotations, mismatched labels, classes imbalance, and model performance to generalize the model is the primary concern.

3. Environmental Variability: Environmental factors should be and more precisely appraised in order to avoid inaccuracies in sensor data quality and difficulties in activity recognition performance. To cater for the differing surroundings, like indoor and outdoor settings, divergent lighting conditions, and diverse background noise one must be able to suit the system to the given environments.

4. User Engagement and Feedback: The introduction of the feedback mechanisms on user engagement and data accuracy technique together with the performance improvement for the activity recognition should be administered. Eliciting user's feedback (input/ preferences/ corrections), which could smooth and refine the system's forecast to produce more accurate and satisfied results.

5. Long-Term Reliability: Providing a longer service life and preventing the system failure after prolonged application. Improving the device performance on the basis of such factors as sensor aging, drifting, model degradation though the means of periodical recalibration, maintenance, and automatic retraining.

6. Ethical Considerations: Since the discussion of the ethical concerns of employing activity recognition systems is wide ranging, it specializes in crucial domains such as healthcare and surveillance. Creating mechanisms to eliminate biases, discrimination and uninfluenced effects in the generalizing process through collecting and managing activity data.

7. Interpretability and Transparency: Improving the process of interpreting and clarification of exercise recognition models to build user trust and explain things. The model output can be explained by describing the reasons for the model's prediction, as well as by marking the most relevant features, which the model uses for making its predictions, and also by providing the users with the opportunity for inspection and validation of the model's behavior.

8. Scalability and Deployment: The system architecture should be engineered to work efficiently in terms of scaling to accommodate big deployment in a complex utility world comprising of heterogeneous users and environments. Improving resource performance, reducing infrastructure cost, and assuring software is operable across many hardware platforms and operating systems. Handling these areas of emphasis will hopefully assemble up a sound, reliable and user-friendly human activity recognition system which is a result of process development that holds the users' interest and meets ethical and privacy standards.

FUTURE DEVELOPMENT PLANS

The potential areas to explore and consider:

1. Enhanced Sensor Fusion: Examine the most sophisticated way of data from various sensors such as wearable devices, cameras, environmental sensors, and IoT; discuss suitable and valuable methods. Consider innovative methodologies of fusing sensor data in order to correct recognition failures, increase robustness of the system and provide contextual information.

2. Advanced Machine Learning Models: Apply the frontline machine learning models, like deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), to uncover various intricate relationships between human motions in both space and time. Examine the use of transfer learning, meta-learning, and self-supervised learning techniques to capitalize on models that have been pre-trained and correcting them accordingly for each activity classification task.

3. Context-Aware Activity Recognition: Equip your algorithms with modules that are capable of distinguishing activities in complex environments and by adding contextual factors to include location, time, social context, and personal preferences. Investigate in the context-aware training methodology that include a graph representation, attention mechanism, and the reinforcement learning.

4. Incremental Learning and Adaptation: Find out concepts which would cause the system to incrementally learn and dynamically adjust to the changing environment as the users' behaviors develops and the system operates. Provide programs for lifelong and meta-learning and ensure self-adaptation mechanism to enable a smooth process of introducing and integrating new tasks and interactions of the users.

5. Edge Computing and Edge AI: Explore the possibility that such activity recognition models shall be deployed on edge computing devices like smartphones, wearables, and edge servers in order to result in the real-time processing and decision-making that is close to the data handling location. Apart from lightweight architecture suppose to push on model compression methods along with on-device model training techniques to achieve efficient model performance and energy consumption on resource-scarce devices.

6. Privacy-Preserving Activity Recognition: Draft privacy-supporting procedures for human recognition, to prevent the privacy and security issues from arising and occurring. Examine federated learning, differential privacy, as well as encryption approaches, designed to permit sharing of knowledge gathered from different data sources without revealing any confidential information.

7. Human-Centric Design and Interaction: Apply user-centered-design (UCD) and iteratively refine the system by including feedback from end-users. Test the system by doing user research, usability analysis, and field trials to learn how the system is used in the real world and how the system meets the needs of the users.

8. Cross-Domain Applications: Thinking about activity recognition technologies that are usable in other fields than their initial applications, e. g. , healthcare, smart homes, office productivity, assisted living and customized services. Network with field experts and involved parties to come up with the applications scenarios along with input and innovation initiatives.

REFRENCES

1. Ahmed Fouad Mohamed Soliman Ali and Kenji Terada. A general framework for multihuman tracking using kalman filter and fast mean shift algorithms. Journal of Universal Computers Science, 16(6):921–937, 2010.
2. Saad Ali and Mubarak Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pages 1–6. IEEE, 2007.
3. Ernesto L Andrade, Scott Blunsden, and Robert B Fisher. Modelling crowd scenes for event detection. In Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, volume 1, pages 175–178. IEEE, 2006.
4. Robert Bodor, Bennett Jackson, and Nikolaos Papanikolopoulos. Vision-based human tracking and activity recognition. In Proc. of the 11th Mediterranean Conf. on Control and Automation, volume 1. Citeseer, 2003.
5. Wongun Choi, Khuram Shahid, and Silvio Savarese. What are they doing?: Collective activity classification using spatio-temporal relationship among people. In Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, pages 1282–1289. IEEE, 2009.
6. Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In In CVPR, pages 886–893, 2005.
7. Charles Dubout and Franois Fleuret. Exact acceleration of linear object detectors. In ECCV, 2012. 7.
8. P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9):1627–1645, 2010.
9. Matteo Frigo, Steven, and G. Johnson. The design and implementation of fftw3. In Proceedings of the IEEE, pages 216–231, 2005.
10. R. B. Girshick, P. F. Felzenszwalb, and D. McAllester. Discriminatively trained deformable part models, release 5. http://people.cs.uchicago.edu/ rbg/latent-release5/.
11. Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th international joint conference on Artificial intelligence, 1981.

12. Subhasis Chaudhuri Neha Bhargava and Guna Seetharaman. Linear cyclic pursuit based prediction of personal space violation in surveillance video. Technical report, Applied Imagery Pattern Recognition Workshop, 2013.
13. Arpita Sinha. Multi-agent consensus using generalized cyclic pursuit strategies. 2009.
14. Viswanthan Srikrishnan and Subhasis Chaudhuri. Crowd motion analysis using linear cyclic pursuit. In Pattern Recognition (ICPR), 2010 20th International Conference on,pages 3340–3343. IEEE, 2010.
15. Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. Technicalreport, International Journal of Computer Vision, 1991.
16. Paul Viola and Michael Jones. Robust real-time object detection. In International Journal of Computer Vision, 2001.
17. Beibei Zhan, Dorothy N Monekosso, Paolo Remagnino,
18. Sergio A Velastin, and Li-Qun Xu. Crowd analysis: a survey. Machine Vision and Applications, 19(5-6):345–357,2008.
19. Vallathan G., John A., Thirumalai C., Mohan S., Srivastava G., Lin J.C.W. Suspicious activity detection using deep learning in secure assisted living IoT environments. J. Supercomput. 2021;77:3242–3260. doi: 10.1007/s11227-020-03387-8. [CrossRef] [Google Scholar]
20. Ullah W., Ullah A., Hussain T., Muhammad K., Heidari A.A., Del Ser J., Baik S.W., de Albuquerque V.H.C. Artificial Intelligence of Things-assisted two-stream neural network for anomaly detection in surveillance Big Video Data. Futur. Gener. Comput. Syst. 2022;129:286–297. doi: 10.1016/j.future.2021.10.033. [CrossRef] [Google Scholar]
21. Doshi K., Yilmaz Y. Rethinking Video Anomaly Detection–A Continual Learning Approach; Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); Waikoloa, HI, USA. 3–8 January 2022; pp. 3036–3045. [Google Scholar]
22. Gumaei A., Hassan M.M., Alelaiwi A., Alsalman H. A Hybrid Deep Learning Model for Human Activity Recognition Using Multimodal Body Sensing Data. IEEE Access. 2019;7:99152–99160. doi: 10.1109/ACCESS.2019.2927134. [CrossRef] [Google Scholar]
23. Uddin M.Z., Hassan M.M. Activity Recognition for Cognitive Assistance Using Body Sensors Data and Deep Convolutional Neural Network. IEEE Sens. J. 2019;19:8413–8419. doi: 10.1109/JSEN.2018.2871203. [CrossRef] [Google Scholar]

24. Taylor W., Shah S.A., Dashtipour K., Zahid A., Abbasi Q.H., Imran M.A. An intelligent non-invasive real-time human activity recognition system for next-generation healthcare. Sensors. 2020;20:2653. doi: 10.3390/s20092653. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
25. Bhattacharya D., Sharma D., Kim W., Ijaz M.F., Singh P.K. Ensem-HAR: An Ensemble Deep Learning Model for Smartphone Sensor-Based Human Activity Recognition for Measurement of Elderly Health Monitoring. Biosensors. 2022;12:393. doi: 10.3390/bios12060393. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
26. Issa M.E., Helmi A.M., Al-Qaness M.A.A., Dahou A., Abd Elaziz M., Damaševičius R. Human Activity Recognition Based on Embedded Sensor Data Fusion for the Internet of Healthcare Things. Healthcare. 2022;10:1084. doi: 10.3390/healthcare10061084. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
27. Hsu S.C., Chuang C.H., Huang C.L., Teng P.R., Lin M.J. A video-based abnormal human behavior detection for psychiatric patient monitoring; Proceedings of the 2018 International Workshop on Advanced Image Technology (IWAIT); Chiang Mai, Thailand. 7–9 January 2018; pp. 1–4. [Google Scholar]
28. Ko K.E., Sim K.B. Deep convolutional framework for abnormal behavior detection in a smart surveillance system. Eng. Appl. Artif. Intell. 2018;67:226–234. doi: 10.1016/j.engappai.2017.10.001. [CrossRef] [Google Scholar]
29. Gunale K.G., Mukherji P. Deep learning with a spatiotemporal descriptor of appearance and motion estimation for video anomaly detection. J. Imaging. 2018;4:79. doi: 10.3390/jimaging4060079. [CrossRef] [Google Scholar]
30. Zhang J., Wu C., Wang Y., Wang P. Detection of abnormal behavior in narrow scene with perspective distortion. Mach. Vis. Appl. 2018;30:987–998. doi: 10.1007/s00138-018- 0970-7. [CrossRef] [Google Scholar]
31. Founta A.M., Chatzakou D., Kourtellis N., Blackburn J., Vakali A., Leontiadis I. A unified deep learning architecture for abuse detection; Proceedings of the WebSci 2019—Proceedings of the 11th ACM Conference on Web Science; Boston, MA, USA. 30 June—3 July 2019; pp. 105–114. [Google Scholar]
32. Moukafih Y., Hafidi H., Ghogho M. Aggressive Driving Detection Using Deep Learning-based Time Series Classification; Proceedings of the 2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA); Sofia, Bulgaria. 3–5 July 2019; pp. 1–5. [Google Scholar]

33. Sabzalian B., Marvi H., Ahmadyfard A. Deep and Sparse features for Anomaly Detection and Localization in video; Proceedings of the 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA); Tehran, Iran. 6–7 March 2019; pp. 173–178. [Google Scholar]
34. Lee J., Shin S. A Study of Video-Based Abnormal Behavior Recognition Model Using Deep Learning. Int. J. Adv. Smart Converg. 2020;9:115–119. [Google Scholar]
35. Bhargava A., Salunkhe G., Bhosale K. A comprehensive study and detection of anomalies for autonomous video surveillance using neuromorphic computing and self learning algorithm; Proceedings of the 2020 International Conference on Convergence to Digital World—Quo Vadis (ICCDW); Mumbai, India. 18–20 February 2020; pp. 2020–2023. [Google Scholar]
36. Xia L., Li Z. A new method of abnormal behavior detection using LSTM network with temporal attention mechanism. J. Supercomput. 2021;77:3223–3241. doi: 10.1007/s11227-020-03391-y. [CrossRef] [Google Scholar]

HAR Report.docx

ORIGINALITY REPORT

10%	8%	5%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	bspace.buid.ac.ae Internet Source	1%
2	dspace.lpu.in:8080 Internet Source	1%
3	Submitted to Liverpool John Moores University Student Paper	1%
4	ijircce.com Internet Source	1%
5	assets.researchsquare.com Internet Source	1%
6	www.mdpi.com Internet Source	<1%
7	Submitted to University of Wales Institute, Cardiff Student Paper	<1%
8	fastercapital.com Internet Source	<1%
9	www.frontiersin.org	

Internet Source

<1%

10 www.journal.esrgroups.org
Internet Source

<1%

11 Yujun Wang, Muhammad Shoaib Saleem, Zahida Perveen, Muhammad Imran. "Jensen, Hermite–Hadamard, and Fejér-Type Inequalities for Reciprocally Strongly (h, s)-Convex Functions", Journal of Mathematics, 2023
Publication

<1%

12 arxiv.org
Internet Source

<1%

13 ijritcc.org
Internet Source

<1%

Huazhen Zhong, Fan Zhang, Yining Zhao, Weifang Zhang, Wenjie Xiao, Xuehai Tang, Liangjun Zang. "UDAD: An Accurate Unsupervised Database Anomaly Detection Method", 2023 IEEE International Performance, Computing, and Communications Conference (IPCCC), 2023
Publication

<1%

Khadija Hicham, Sara Laghmati, Amal Tmiri. "3D CT Scans for Colorectal Cancer Classification using VGG16 and Data Augmentation", 2023 10th International

<1%

Conference on Wireless Networks and Mobile Communications (WINCOM), 2023
Publication

16	Submitted to University of Glamorgan Student Paper	<1%
17	Submitted to Glasgow Caledonian University Student Paper	<1%
18	sitecproject.com Internet Source	<1%
19	www.geeksforgeeks.org Internet Source	<1%
20	João Monge, Gonçalo Ribeiro, António Raimundo, Octavian Postolache, Joel Santos. "AI-Based Smart Sensing and AR for Gait Rehabilitation Assessment", Information, 2023 Publication	<1%
21	Ton Duc Thang University Publication	<1%
22	Tran Duc Luong, Vuong Minh Tien, Phan The Duy, Van-Hau Pham. "Chapter 2 FedLS: An Anti-poisoning Attack Mechanism forFederated Network Intrusion Detection Systems Using Autoencoder-Based Latent Space Representations", Springer Science and Business Media LLC, 2023 Publication	<1%

23	mau.diva-portal.org Internet Source	<1%
24	medium.com Internet Source	<1%
25	uplatz.com Internet Source	<1%
26	K. Yin, J. Jia, F. Li, X. Gao, T. Sun. "Harnessing edge-enhanced attention mechanisms for supernova detection in deep learning frameworks", Astronomy and Computing, 2024 Publication	<1%
27	Lecture Notes in Computer Science, 2011. Publication	<1%
28	Md Tanzil Shahria, Jawhar Ghommam, Raouf Fareh, Mohammad Habibur Rahman. "Vision-Based Object Manipulation for Activities of Daily Living Assistance Using Assistive Robot", Automation, 2024 Publication	<1%
29	Submitted to University of Strathclyde Student Paper	<1%
30	dataanalysisexperts.net Internet Source	<1%
31	technodocbox.com Internet Source	<1%

32	www.analyticsvidhya.com Internet Source	<1%
33	www.ijraset.com Internet Source	<1%
34	www.medrxiv.org Internet Source	<1%
35	www.semanticscholar.org Internet Source	<1%

Exclude quotes	On	Exclude matches	< 3 words
Exclude bibliography	On		