**Integrated Motion Analysis of Humans**

A Project Report submitted in partial fulfillment of the requirements for the award of the degree of

**Bachelor of Technology**

### in

**Computer Science and Engineering and Electronics and Communication Engineering**

by

### Ish Savaliya (112115142)

### Shrunmay Shinde (112115148)

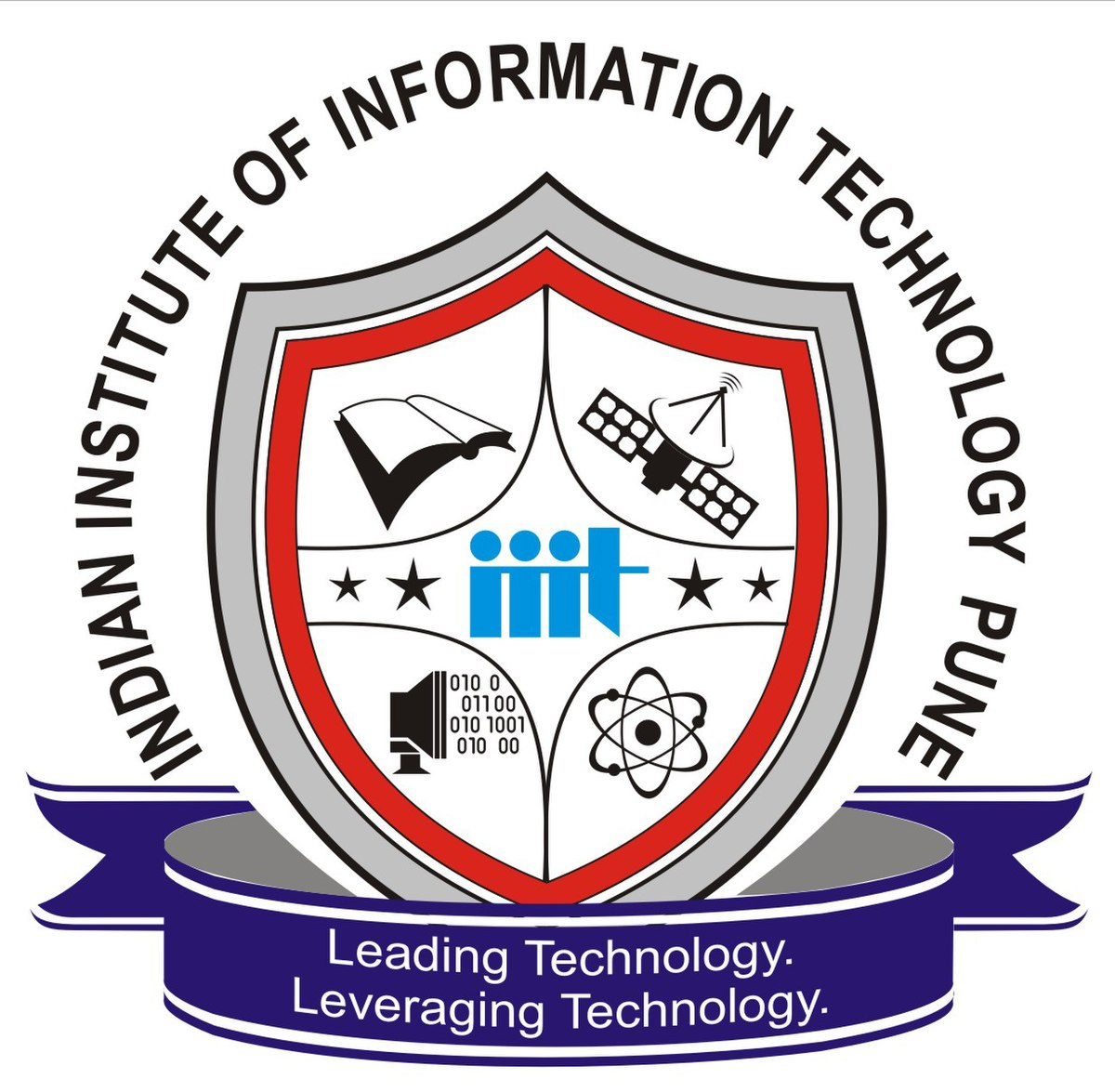
### Aditya Bhusari (112116002)

### Daksh Agarwal (112116012)

### Tejas Patil (112116042)

**Under the Supervision of: Dr. Meenakshi Choudhary**

### Semester: Fifth

****

#### Department of Computer Science and Engineering &

#### Department of Electronics and Communication Engineering

#### 

#### Indian Institute of Information And Technology, Pune

**(An Institute of National Importance by an Act of Parliament)**

#### November 2023

**BONAFIDE CERTIFICATE**

This is to certify that the project report entitled **“Integrated Motion Analysis of Humans”** submitted by **Ish Savaliya** bearing the **MIS No: 112115142**, **Shrunmay Shinde** bearing the **MIS No: 112115148**, **Aditya Bhusari** bearing the **MIS No: 112116002**, **Daksh Agarwal** bearing the **MIS No: 112116012, Tejas Patil** bearing the **MIS No: 112116042** in completion of their project work under the guidance of **Dr. Meenakshi Choudhary** is accepted for the project report submission in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in the **Department of Computer Science and Engineering and Electronics and Communication Engineering**, Indian Institute of Information Technology, Pune (IIIT Pune), during the academic year **2023-24**.

#### Dr. Meenakshi Choudhary Dr. Sanjeev Sharma Dr. Sushant Kumar

Project Guide Head of the Department Head of the Department

Assistant Professor Assistant Professor Assistant Professor

Computer Science and Engineering Department of CSE Department of ECE

IIIT Pune IIIT Pune IIIT Pune

Project Viva-voce held on: 06/11/2023

**Undertaking for Plagiarism**

We **Ish Savaliya, Shrunmay Shinde, Aditya Bhusari, Daksh Agarwal, Tejas Patil** solemnly declare that research work presented in the **report/dissertation** titled **“Integrated Motion Analysis of Humans”** is solely **our** research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete report/dissertation has been written by **us**. I understand the zero tolerance policy of **Indian Institute of Information Technology Pune** towards plagiarism. Therefore **we** declare that no portion of my **report/dissertation** has been plagiarized and any material used as reference is properly referred/cited. I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of the degree, the Institute reserves the rights to withdraw/revoke my **B.Tech** degree.

**Students’/ Student’s Name and Signature with Date**

**Conflict of Interest**

**Manuscript title: Integrated Motion Analysis of Humans**

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers’ bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

**Students’/ Student’s Name and Signature with Date**

## ACKNOWLEDGEMENT

This project would not have been possible without the help and cooperation of many. I would like to thank the people who helped me directly and indirectly in the completion of this project work.

First and foremost, I would like to express my gratitude to our honorable Director, **Prof. O.G. Kakde**, for providing his kind support in various aspects. I would like to express my gratitude to my project guide **Dr. Meenakshi Choudhary**, **Department of CSE**, for providing excellent guidance, encouragement, inspiration, constant and timely support throughout this **B.Tech Project**. I would like to express my gratitude to the **Dr. Sanjeev Sharma,** **Head of Department**, **Department of CSE** and **Dr. Sushant Kumar, Head of Department, Department of ECE**, for providing his kind support in various aspects. I would also like to thank all the faculty members in the **Department of CSE/ECE** and my classmates for their steadfast and strong support and engagement with this project.

**Integrated Motion Analysis of Humans**

## Abstract

The deployment of surveillance systems in public places has become increasingly pervasive, aiming to enhance security and monitor public spaces. This explores the multifaceted landscape of surveillance in public areas, with a focus on the delicate balance between security and privacy.

In this project, we delve into the intricate realm of small crowd dynamics, aiming to decode collective behavior within limited gatherings. Utilizing the fusion of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, we explore subtle patterns defining social interactions in confined spaces. Our study is rooted in the diverse UCF50 dataset, representing various small group scenarios. By merging spatial awareness and temporal understanding, we decipher the unique language of small crowd behavior. Our insights hold implications for event planning, retail analytics, and public safety. This project advances computer vision techniques, offering a profound understanding of group dynamics within small crowds.

**Keywords:** Small crowd dynamics, CNN-LSTM networks, UCF50 dataset , retail analytics, computer vision advancements.

## TABLE OF CONTENTS

**Abstract 1**

**(i)** [**List of Figures/Symbols/Nomenclature 2**](#_heading=h.gjdgxs)

**(ii)** [**List of Tables 3**](#_heading=h.30j0zll)

1. **Introduction 4**

[1.1 Overview of work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4](#_heading=h.1fob9te)

[1.2 Motivation of work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4](#_heading=h.3znysh7)

[1.3 Literature Review . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5](#_heading=h.3znysh7)

1.4 Research Gap. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

1. **Problem Statement**

2[.1 Research Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7](#_heading=h.3znysh7)

2[.2 Methodology of work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8](#_heading=h.3znysh7)

1. **Analysis And Design 11**
2. **Results and Discussion 17**
3. **Conclusion and Future Scope 20**

**6 References**

# 

# List of Figures

# Figure1] Task 1

# Figure2] Task 2

# Figure3] Task 3

# Figure4] Task 4

# Figure5] Project Prototype

# Figure6] LRCN Model

# Figure7] LRCN Model

# Figure8] Architecture of LRCN Model

# Figure9] Homepage Interface

# Figure10] Training Accuracy vs Validation Accuracy

# Figure11] Training Loss vs Validation Loss

# Figure12] Display Interface of Predicted Result

# List of Tables

# Table 1]: Comparison of Proposed Model with the existing models mentioned in Research papers discussed in literature review.

**Chapter 1**

**Introduction**

## Overview of Work

Our LRCN model represents a cutting-edge blend of computer vision and deep learning, specifically designed to unravel complex social interactions within small groups. Leveraging the capabilities of Long-term Recurrent Convolutional Networks (LRCNs), this model seamlessly integrates spatial analysis and temporal patterns, enabling a detailed comprehension of human behavior. Trained on the diverse UCF50 dataset, LRCN accurately identifies activities like "Bench Press," "Fencing," and more, providing crucial insights into group dynamics. With applications spanning event management, retail analytics, and public safety, this model redefines our understanding of social interactions in confined spaces. Additionally, when coupled with Flask, it offers real-time processing, enhancing its usability and making it an indispensable tool for various domains.

## Motivation of the Work

Our project is motivated by the challenge of understanding social interactions within small groups. We aim to unravel the complexities of how people in limited gatherings behave and interact. By using advanced technology, specifically our LRCN model, we delve into the intricate patterns that define these group dynamics. Our curiosity is fueled by the diverse UCF50 dataset, which captures various actions within small groups. This dataset is the foundation of our exploration. We believe that by decoding these subtle social cues, we can gain valuable insights. These insights not only help us understand human behavior but also have practical applications in areas like event planning, retail analysis, and public safety. Through our research, we want to bridge the gap between technology and real-world social interactions. By applying the power of LRCN, we hope to contribute to creating more connected and responsive communities. Our goal is to make a positive impact on society by deciphering the language of small group dynamics, making interactions smoother, and fostering a deeper understanding among people.

[**1.3 Literature Review**](#_heading=h.3znysh7)

**[1]** The 3D CNN architecture is used for video classification by utilizing spatial temporal features . The architecture consists of six 3D convolutional layers and 4 max pooling 3D layers. It creates a 3D activation map and further the 3D filter is used to calculate the representation of elements at a low level. The UCF Youtube Action Dataset is used and overall accuracy of 85.2% is obtained.

**[2]** The architecture is build using CNN and LSTM . CNN is used to extract spatial features and LSTM for extracting temporal features. A new dataset is developed using Kinect V2 sensor and 12 different classes of categories. The movements of body part is plotted on a graph and the model is trained using the x and y axis coordinates in csv file. The average accuracy obtained is 87%.

**[3]** The architecture used is CNN + SVM on dataset UCF101. Used 2 stream 2D CNN that is trained on both spatial and temporal domains of video data. The temporal domain takes stacked multi frame dense optical flow data as input. The major weakness is that the huge storage is required even for small dataset and the accuracy obtained is 88%. The performance is worst in temporal domain of video.

**[4]** The ConvLSTM model uses the raw skeletal coordinates along with their characteristic geometrical and kinematic features to construct the novel guided features. KinectHAR dataset is used collected with the help of KinnectV2 sensor. The dataset containing joint values alongwith extracted manual features are inputed to deep learning network i.e CNN, LSTM for activity recognition and fall detection. The F1 score obtained is 90.40%.

**[5]** Various architectures like CNN, BiLSTM etc. are implemented on PAMAP2 dataset and testing accuracy obtained is 91.00% and 89.52% respectively.

**[6]** A novel mechanism that utilizes CNN free approach and then using sequential learning method to achieve the state of an art accuracy of 73.714% on HMDB51 dataset. Multi layered LSTM is used to learn temporal relationships among these features to recognize human activities with higher precision. Multi layered LSTM gives accuracy of 72.1% on HMDB51 dataset.

**1.4 Research Gap**

Each of the model implemented in papers presented in literature review has its disadvantages.

* The model used in [1] have used more number of convolutional blocks as compared to our model but haven’t achieved the accuracy better than us. Also the model is highly expensive for computation.
* The hybrid model in [2] shows excellent performance on activity recognition of one- person activity, but don’t perform good in case of multiple people.
* The model implemented in [3] requires huge storage requirement even for a small dataset and then also the accuracy is not as expected.
* The model used in [4] have limitations in terms of computational complexity and there is a need of powerful hardware.
* The model used in [5] have limitations in capturing fine grained spatial details. Limitations in ability to handle complex spatial information. Overfitting limits the model’s ability to generalize the unseen data.
* Model used in [6] is computationally complex, requiring significant resources, which may not be suitable for resource constrained wearable devices. No well defined method for determining the size of the sample data is given.

# Chapter 2

# Problem Statement

Our challenge is to create a unified system that combines computer vision and machine learning to interpret human movements accurately. Existing methods are limited, focusing on specific aspects and missing the broader context. This gap hinders progress in fields like healthcare, security, sports, and psychology, where understanding human actions is crucial. We need a solution that can analyze human actions, providing valuable insights without complexity. Our goal is to develop an integrated motion analysis system that seamlessly interprets human movements. By merging computer vision and machine learning, our aim is to revolutionize various sectors by offering straightforward and actionable insights into human behavior.

**2.1. Research Objectives**

The main objective behind our model, which combines LRCN (Long-term Recurrent Convolutional Network) with Flask, is to create a user-friendly system that can recognize and categorize human actions within small groups accurately. By using LRCN, we enable the model to understand both the spatial and temporal aspects of these actions, making the recognition more precise. Integrating Flask allows us to deploy this model on a web platform, making it accessible to users. Our aim is to provide a seamless experience for individuals interacting with the system. Users can upload videos of small group activities, and our LRCN model, powered by Flask, processes these videos. The system then identifies and categorizes the actions, displaying the results instantly. This setup ensures that our model is not only accurate but also user-friendly, enabling effortless interaction and understanding of group behaviors.

**2.2 Methodology of the Work:**

1. **Neural Network model implemented:** ConvLSTM2D (A model that combines the convolutional layers with Long Short Term Memory (LSTM) recurrent neural network).

2. **Dataset Used:** UCF-50 Action Recognition Dataset sourced from Kaggle. Dataset used consists of 50 action categories. Each category consists of 150 videos approximately of length 15 seconds each in .avi format.

Dataset Source: <https://www.kaggle.com/datasets/vineethakkinapalli/ucf50-action-recognition-dataset>

3. **Libraries Used:**

i. **TensorFlow**: A open-source library used for data preprocessing, model implementation and for using some callback functions.

ii. **Computer Vision (cv2)**: A open-source library used for video preprocessing and frame extraction from videos.

iii. **Matplotlib:** A open-source library used for visualization of performance metrics.

4. **Data Preprocessing:**

**Task 1:** Extracting the frames from videos and resized to predefined shape of 64 x 64 pixels. This frames are then normalized and appended in list for each video. The frames are considered at regular interval of 20 frames.

**Task 2:** Create a dataset which consisting of three attributes namely **features, labels, video file paths**. The features is an numpy array consisting of list of frames for a video made in first task and labels is also an numpy array consisting of labels associated with each features. The video path is a list containing the path of video.

**Task 3:** Encoding labels using one hot encoder.

**Task 4:** The last task is to split the dataset into train and test dataset.

The tasks mentioned above can be visualized as follows:

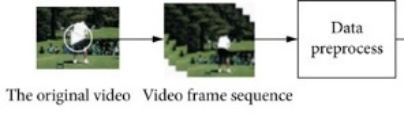


Figure1]: Task 1

Source: Google Images

Figure3]: Task 3

0

0

0

0

0

1

0

0

0

0

1

0

Bench Clean And Fencing Hula Hoop Jump Rope Jumping

Press Jerk Jack

Figure2]: Task 2

0

0

0

1

0

0

0

0

1

0

0

0

0

1

0

0

0

0

1

0

0

0

0

0

Jumping Jack

Jump Rope

Hula Hoop

Fencing

Clean And Jerk

Bench Press

Video File

Path

Associated Label

List of Video Frames

Video File Paths

Labels

Features

**Dataset**

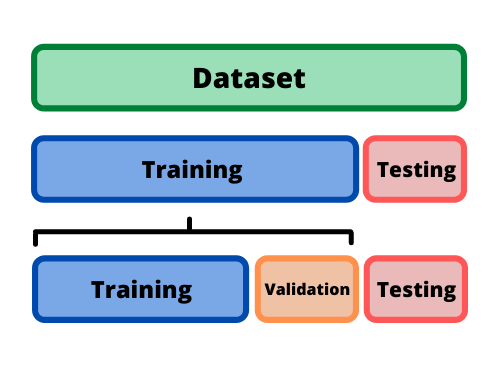
****

Figure4]: Task 4

Source: Google Images

5. **Developing our model:**

Developing our model using various time distributed convolutional layers and max pooling layers along with some dropout layers(25%) and at last adding a Long- Short Term Memory (LSTM) , a recurrent neural network to make our own customized ConvLSTM2d model. All layers added are **‘TimeDistributed’** layers to deal with sequence of data. ‘**TimeDistributed’** is a wrapper layer that allows to apply a layer, typically a dense (fully connected layer) to each time step of sequence independently.

6. **Model Training and Testing:**

i. Compilation of model using **Adam** as optimizer, **Categorical Cross Entropy** as loss function and **Accuracy** as metrics.

ii. Using **Early Stopping** callback function to stop training if validation loss is not decreasing with a patience of 20.

iii. Training the model with **70 epochs** with batch size 4 and steps per epochs are 117. Validation split is 20%.

iv. Evaluating model over the test set.

7. **Saving the model for feeding it to backend for prediction of unseen videos.**

8. **Plotting Accuracy vs Validation Accuracy and Loss vs Validation Loss graphs.**

# Chapter 3

# Analysis and Design

**Prototype:**

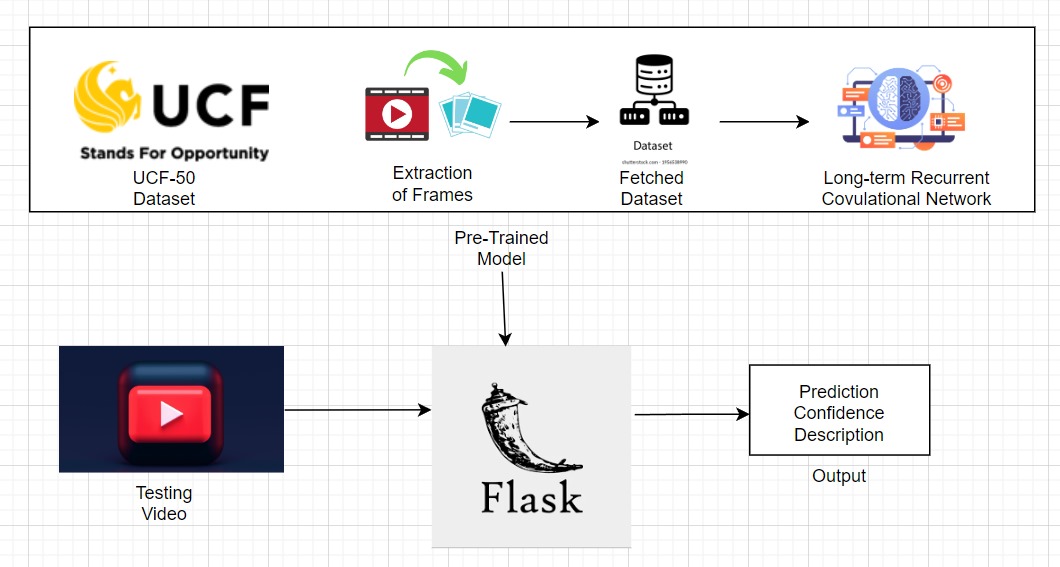
****

Figure5]: Project Prototype

**The layers used are as follows:**

* Conv2D: A Conv2D layer, short for Convolutional 2D layer, is a fundamental building block in convolutional neural networks (CNNs). It is commonly used for processing 2D grid-like data, such as images, but can also be applied to other types of data with grid-like structures, like spectrograms or any 2D grid of values. The primary purpose of a Conv2D layer is to perform convolution operations on the input data. Convolution is a mathematical operation that involves sliding a small filter (also called a kernel) over the input data and computing the element-wise dot product between the filter and the input at each position. This operation allows the network to learn features from the input data, such as edges, textures, and higher-level patterns, by applying multiple filters.
* MaxPooling2D: Max pooling is a common operation used in convolutional neural networks (CNNs) to downsample the spatial dimensions of feature maps produced by convolutional layers. A MaxPooling2D layer, specifically, is used for 2D data, such as images or feature maps, to reduce the spatial resolution while retaining important features.The basic idea of max pooling is to divide the input feature map into small, non-overlapping regions (often 2x2 or 3x3) and for each region, select the maximum value. This maximum value is then used to represent that region in the output.
* TimeDistributed Layer: A TimeDistributed layer is a wrapper or decorator used in recurrent neural networks (RNNs) and sequential models to apply a particular layer to each time step of a sequence independently. This layer is particularly useful when you have sequences of varying lengths, such as in natural language processing (NLP) tasks or time series analysis, where you want to apply a specific layer to each time step of the input sequence. The TimeDistributed layer allows you to wrap another layer (e.g., a Dense layer, LSTM layer, or Conv1D layer) and apply it to each time step of the input sequence independently. This means that the weights of the wrapped layer are shared across all time steps. It's as if you have a separate instance of the layer for each time step, but they all share the same weights.
* Dropout Layer: Dropout is a regularization technique used in neural networks to prevent overfitting. It is a simple but effective method for improving the generalization of a neural network by reducing the risk of the network relying too heavily on any one specific neuron.
* Flatten Layer: A Flatten layer is a layer commonly used in neural networks, particularly in convolutional neural networks (CNNs) and sequential models, to transform multi-dimensional data into a one-dimensional vector. The primary purpose of a Flatten layer is to reshape the input data, making it suitable for fully connected (dense) layers or other layers that expect one-dimensional input.

**Design of our implemented model:**

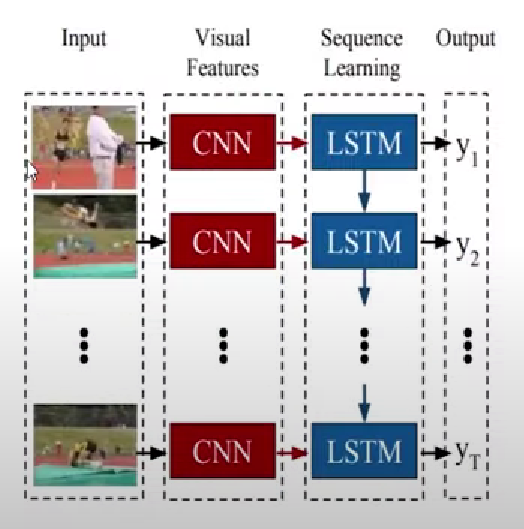
****

Figure6]: LRCN Model

Source: Google Images

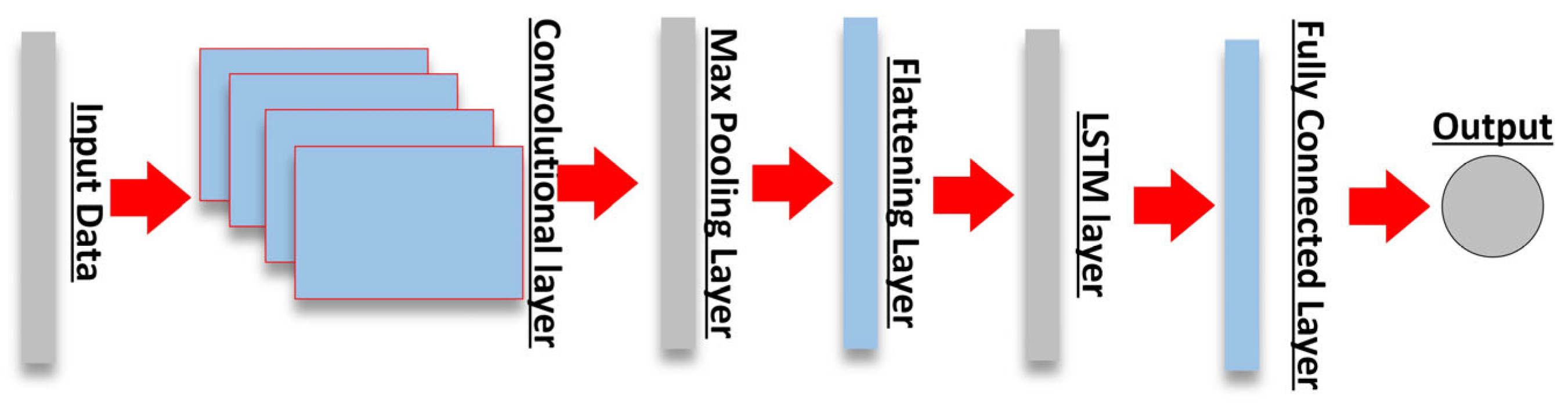


Figure7]: LRCN Model

Source: [8] **MDPI and ACS Style**

# Architecture of our proposed model:

# 

# Figure8]: Architecture of LRCN Model

# Information about Flask used for Backend and Frontend:

# Flask is a Python-based web framework renowned for its simplicity and flexibility, making it an ideal choice for projects that don't require complex dependencies. This micro-framework provides a streamlined approach to web application development and empowers developers to craft applications with precision. Some of the standout features of Flask include:

# Routing: Flask makes it easy to define URL routes and map them to specific Python functions. This enables you to create distinct endpoints for your web application, ensuring that different parts of your site or API can be accessed and interacted with through well-structured URLs.

# Templating: Flask supports template rendering, allowing you to generate HTML content dynamically. By embedding Python code within HTML templates, you can create web pages that adapt to user input and other variables, providing a dynamic and personalized user experience.

# Web Server Integration: Flask is versatile when it comes to web server compatibility. It can run on various web servers, including its built-in development server for local testing, as well as popular options like Gunicorn, uWSGI, and Apache. This flexibility ensures your Flask application can be deployed in various hosting environments.

# Request and Response Handling: Flask simplifies the handling of HTTP requests and responses. It provides convenient methods for accessing data from incoming requests, such as form data and query parameters. Likewise, it offers efficient tools to generate HTTP responses, which are crucial for sending content and information back to clients.

# Extensions: Flask's modular design allows developers to enhance their applications by integrating a wide array of extensions. These extensions cover a broad spectrum of functionalities, including database integration, user authentication, and more. You can cherry-pick the extensions that suit your project's requirements, customizing your application's capabilities as needed.

# Lightweight and Unopinionated: Flask is often referred to as a "micro" framework because it doesn't enforce a specific project structure or include an abundance of built-in features. This lightweight approach empowers developers with the freedom to structure their applications as they see fit, making Flask an excellent choice for a wide range of projects.

# Vibrant Community: Flask boasts a thriving and active community of developers. This means you'll find a wealth of resources, tutorials, and extensions readily available. The support of this community can significantly ease the development process and help you overcome challenges as you build your application.

# Working of Web Application:

# The provided code outlines a Flask web application designed to process and analyze uploaded videos using a pre-trained machine learning model. It begins by importing essential Python libraries such as Flask for web development, TensorFlow for machine learning, OpenCV for video processing, and NumPy for numerical operations.

# The Flask application, denoted as `app`, is created to serve as the foundation of the web service. It loads a pre-trained machine learning model from an HDF5 file and configures the upload folder and acceptable video file extensions, including mp4, avi, and mov.

# Two functions play vital roles in the application's functionality. The `allowed\_file` function checks if an uploaded file has a valid extension. The `process\_video\_with\_model` function handles video analysis by extracting frames, resizing, normalizing them, and utilizing the machine learning model for content predictions.

# The application defines two routes: `/home` serves as the homepage and `/upload` handles video uploads. When a user uploads a video, the `upload\_video` function checks for validity, saves the video, and processes it with the `process\_video\_with\_model` function. The results are then rendered on the web page.

# In conclusion, this code creates a Flask web application that empowers users to upload videos for content analysis using a pre-trained machine learning model. It offers a user-friendly interface for processing video files with a focus on simplicity and flexibility. Debugging is enabled for easier development and testing

# Home Page Interface :

# 

# Figure9]: Home Page Interface

# Chapter 4

# Results and Discussion

# The Model stopped training at the 50th epoch due to early stopping callback function implemented and the validation loss was not improving.

# The training accuracy and validation accuracy obtained in last epoch was 100% and 90.60% respectively.

# The testing accuracy obtained by evaluating the model was 91.28%.

# The Training accuracy vs Validation accuracy graph obtained is:

# 

Figure10]: Training Accuracy vs Validation Accuracy

* The Training loss vs Validation Loss graph obtained is:

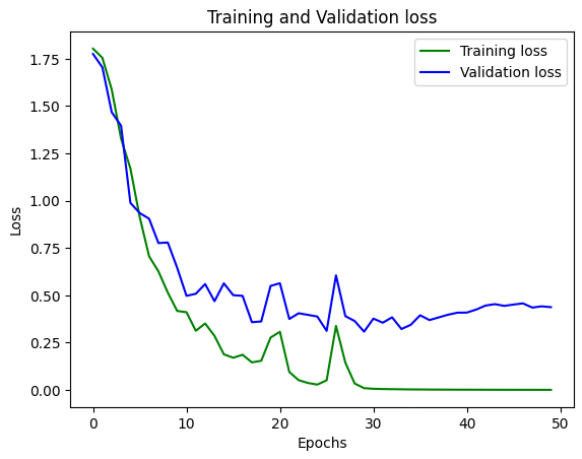


Figure 11]: Training Loss and Validation Loss

**Comparison with existing models as discussed in research papers mentioned in Literature Review**

|  |  |
| --- | --- |
| **Models** | **Accuracy** |
| 3D CNN | 74.2% |
| CNN+LSTM | 88.89% |
| CNN+SVM | 88% |
| ConvLSTM | 90.40% |
| BiLSTM | 89.52% |
| Multi Layered LSTM | 72.10% |
| **LRCN (Proposed Model)** | **91.28%** |

Table1]

**Display Interface of predicted result:**

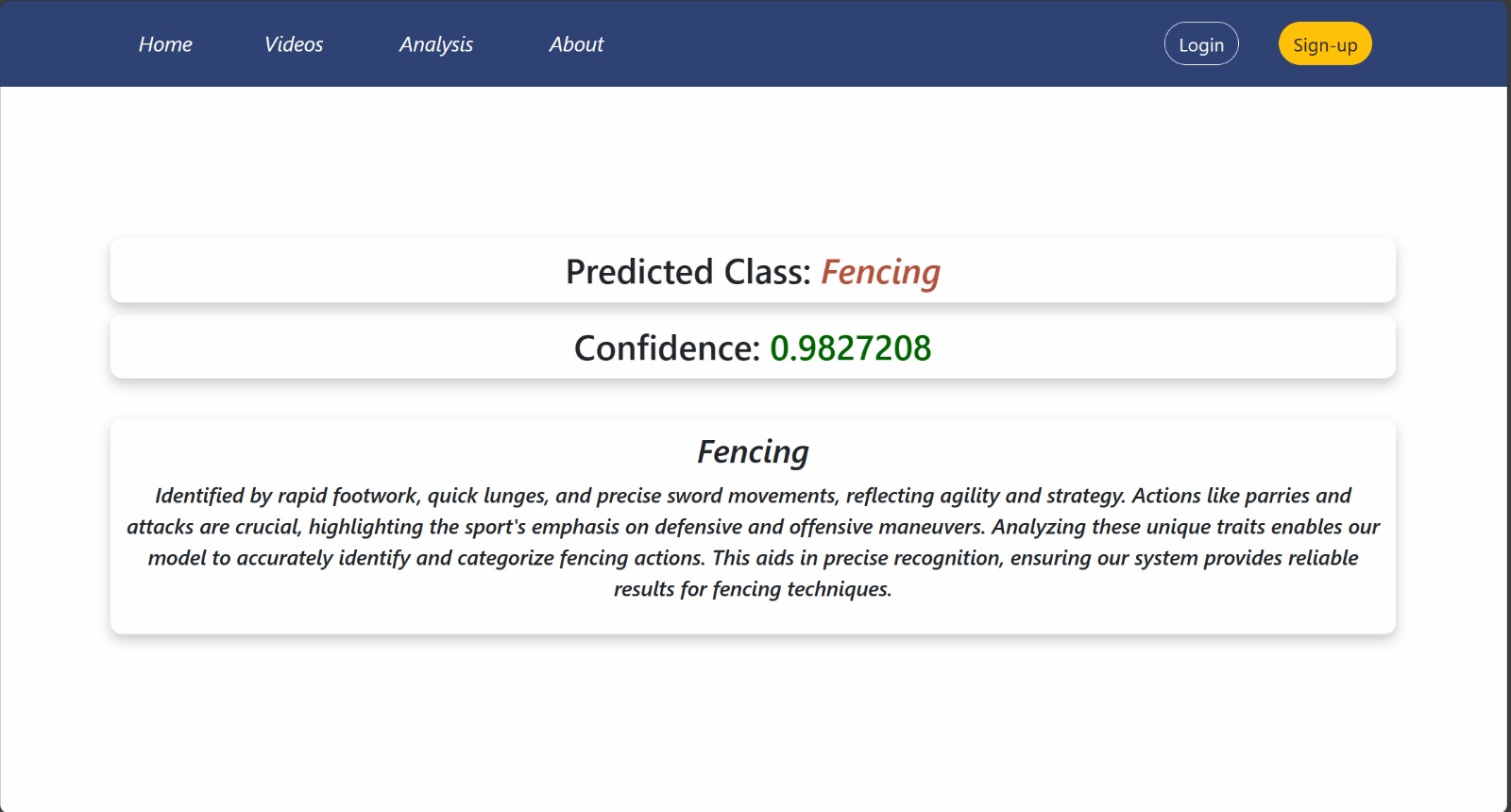
**e**

Figure12]: Display Interface of Predicted Result

# Chapter 5

# Conclusion and Future Scope

**Conclusion:**

* The Integrated Motion Analysis model represents a breakthrough in understanding human actions using technology. By using Long Term Recurrent Convolutional Network (LRCN) and Flask, the model can accurately interpret how people move.
* This model has many practical uses. In security, it can detect suspicious activities, making public spaces safer. In healthcare, it helps monitor patients and track their progress in therapy. It also finds applications in smart homes, sports analysis, and entertainment, improving user experiences in these areas.
* One of its strengths is its flexibility. It can be used in education, retail, and crowd management, showing its versatility. As technology advances, the model can handle imbalanced data better and predict future activities more accurately. It can also adapt to different situations, making it robust and suitable for various environments.
* In the future, the model can integrate with new technologies, making it even better at understanding human behavior. It can become more intuitive, allowing people to interact with computers more naturally. It could also become more user-friendly by incorporating voice recognition and augmented reality features.
* In conclusion, the Integrated Motion Analysis model is a significant step forward in technology. It helps machines understand human actions, making it useful in many fields. Its ability to adapt and improve in the future makes it a valuable tool, enhancing our understanding of human behavior in various real-life situations.

**Future Scope:**

* Analyzing Large Crowd’s Behavior: Extend the scope to understand large group behaviors. Analyzing how people behave in crowds is essential for managing events and public spaces effectively.
* Predicting Future Actions: Combine activity recognition with technologies like Brain-Computer Interfaces to predict what people might do next. This has applications in areas such as driver behavior prediction.
* Handling Environmental Changes: Create systems that work effectively regardless of changes in lighting, weather, or other environmental factors. Robust systems can handle these variations.
* Future Hardware Enhancements: Upgrading video capture with high-resolution cameras, depth-sensing technology, thermal imaging, and wearables provides diverse data sources. Pairing these advancements with powerful processors and efficient data transmission ensures real-time, accurate human behavior analysis. Energy-efficient designs enhance sustainability, promising a comprehensive approach to future applications.
* Integration with Wearable Devices: Extend crowd analysis to wearable devices, enabling event organizers to monitor attendee movements and interactions. This data can enhance event planning, ensuring efficient crowd management and a seamless attendee experience.
* Creating an video archive, and introducing a feedback feature are crucial steps in enhancing web application. These additions will ensure user data security, provide a history of analyzed videos, and allow users to contribute feedback, ultimately enabling continuous improvement of model accuracy and functionality.

# References

[1] Vrskova, R.; Hudec, R.; Kamencay, P.; Sykora, P. “Human Activity Classification Using the 3DCNN Architecture.” Appl. Sci. 2022,12, 931. Date of Publication: 17/01/2022 https://doi.org/10.3390/app12020931 .

[2] Khan, I.U.; Afzal, S.; Lee, J.W. “Human Activity Recognition via Hybrid Deep Learning Based Model.” Sensors 2022, 22, 323. Date of Publication: 01/01/2022 <https://doi.org/10.3390/s22010323>

[3] Md. Milon Islam, Sheikh Nooruddin, Fakhri Karray and Ghulam Muhammad “Human Activity Recognition using tools of convolutional neural networks: A state of art review, datasets, Challenges and future Prospects” Journal: Elsevier. Date of Publication: 01/10/2022

<https://doi.org/10.1016/j.compbiomed.2022.106060>

[4] Yadav, S.K., Tiwari, K., Pandey, H.M. et al. Skeleton-based human activity recognition using ConvLSTM and guided feature learning. Soft Comput 26, 877–890 (2022). Date of Publication: 01/10/2021.

<https://doi.org/10.1007/s00500-021-06238-7>

[5] Zhang, Shibo, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng, and Nabil Alshurafa. 2022. "Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances" Sensors 22, no. 4: 1476. Date of Publication: 14/02/2022. <https://doi.org/10.3390/s22041476>

[6] Altaf Hussain , Tanveer Hussain, Waseem Ullah ,and Sung Wook Baik ,Sejong University, Seoul 143-747, Republic of Korea. “Video Transformer and Deep Sequence Learning for Human Activity Recognition in Surveillance of Videos” Hindawi Publication. Date of Publication: 04/04/2022.

<https://doi.org/10.1155/2022/3454167>

[7] L. Sun, J. Du, L. -R. Dai and C. -H. Lee, "Multiple-target deep learning for LSTM-RNN based speech enhancement," 2017 Hands-free Speech Communications and Microphone Arrays (HSCMA), San Francisco, CA, USA, 2017, pp. 136-140,

doi: 10.1109/HSCMA.2017.7895577.

[8] Aksan, F.; Li, Y.; Suresh, V.; Janik, P. CNN-LSTM vs. LSTM-CNN to Predict Power Flow Direction: A Case Study of the High-Voltage Subnet of Northeast Germany. *Sensors* **2023**, *23*, 901. Published on www.mdpi.com.