Human Activity Recognition using Deep Learning Approach

Submitted in partial fulfillment of the requirements of the degree of

Bachelor of Science in Computer Science and Engineering

Developed By

Habibullah	18192103080
Pallab Majumdar	18192103050
Mafuja Akter Mitu	18192103068
Joy Adhikary	18192103062
Al Ahad Sufian	18192103056

Supervised By

M. M. Fazle Rabbi

Assistant Professor

Dept. of CSE, BUBT

Bangladesh University of Business and Technology - BUBT

May 2023

These categories are showing the classification of this video. We have analyzed the categories of this video.

Figure 3.2: Walking category

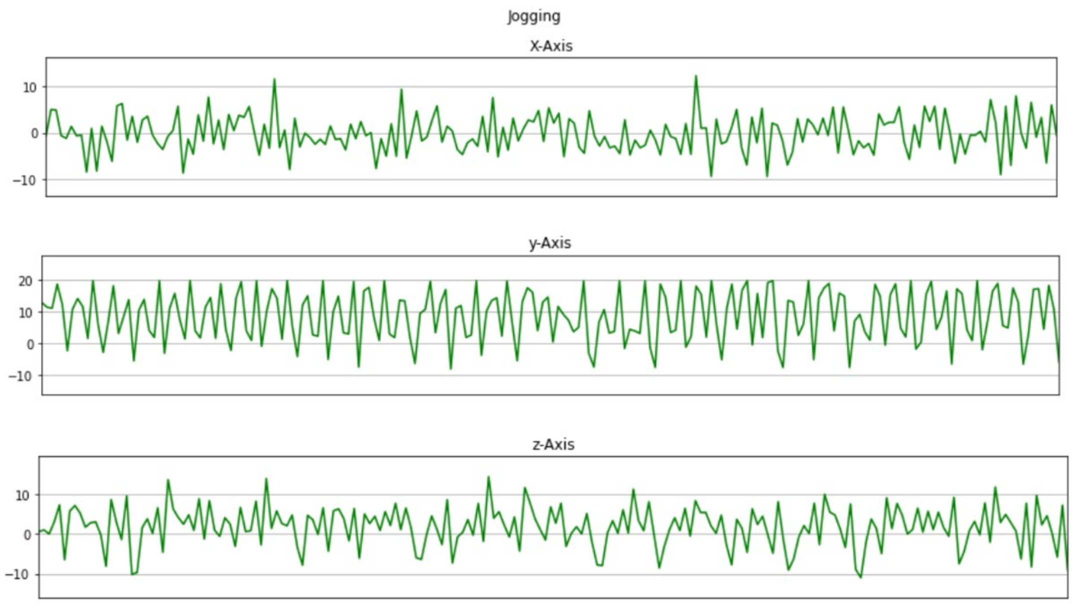


Figure 3.3: Jogging category

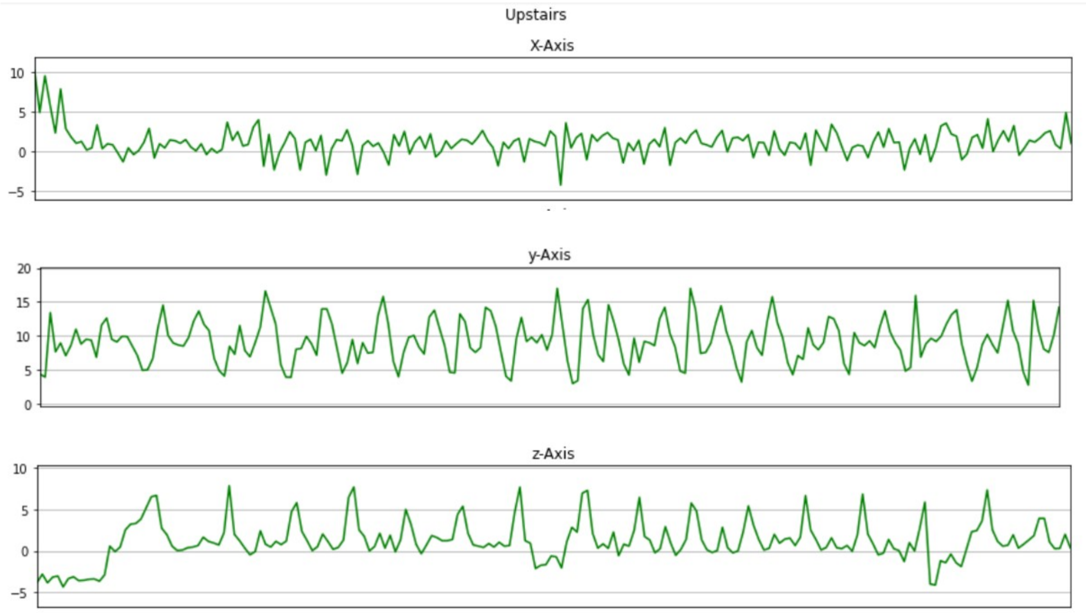


Figure 3.4: Upstairs category

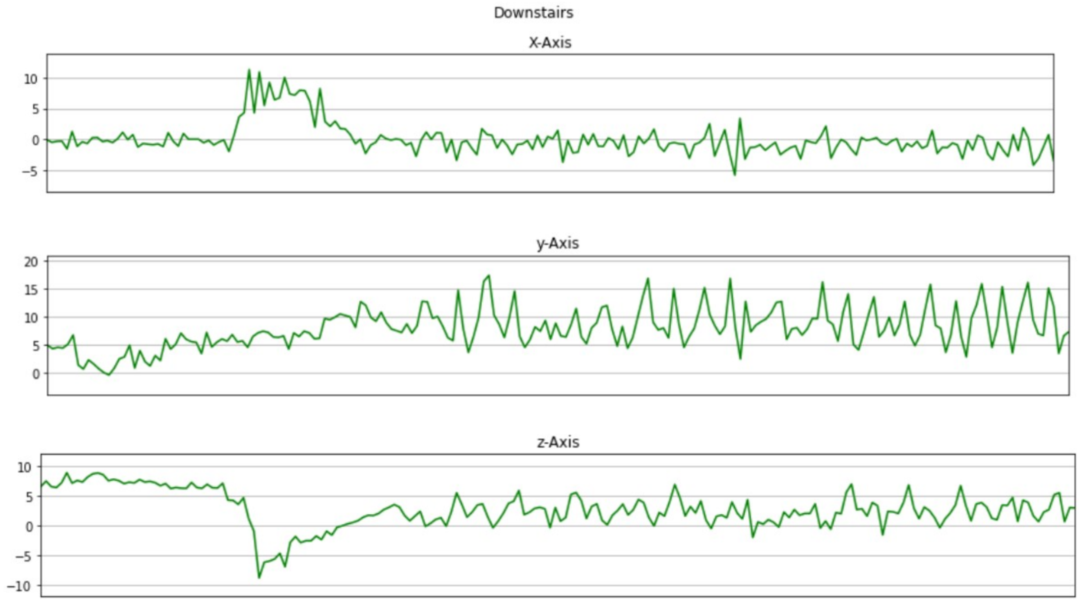


Figure 3.5: Downstairs category

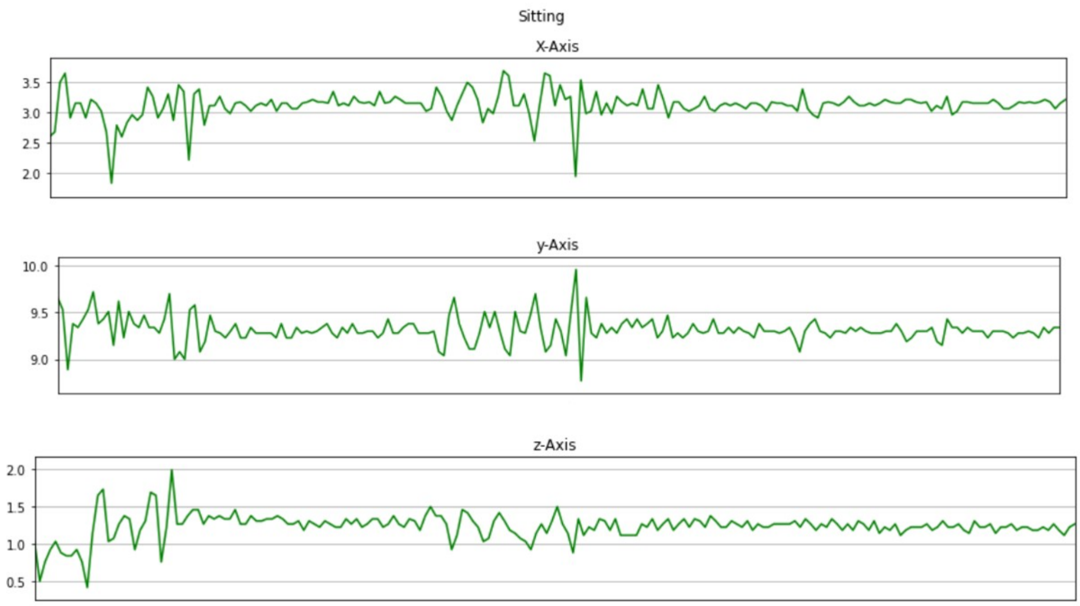


Figure 3.6: Sitting category

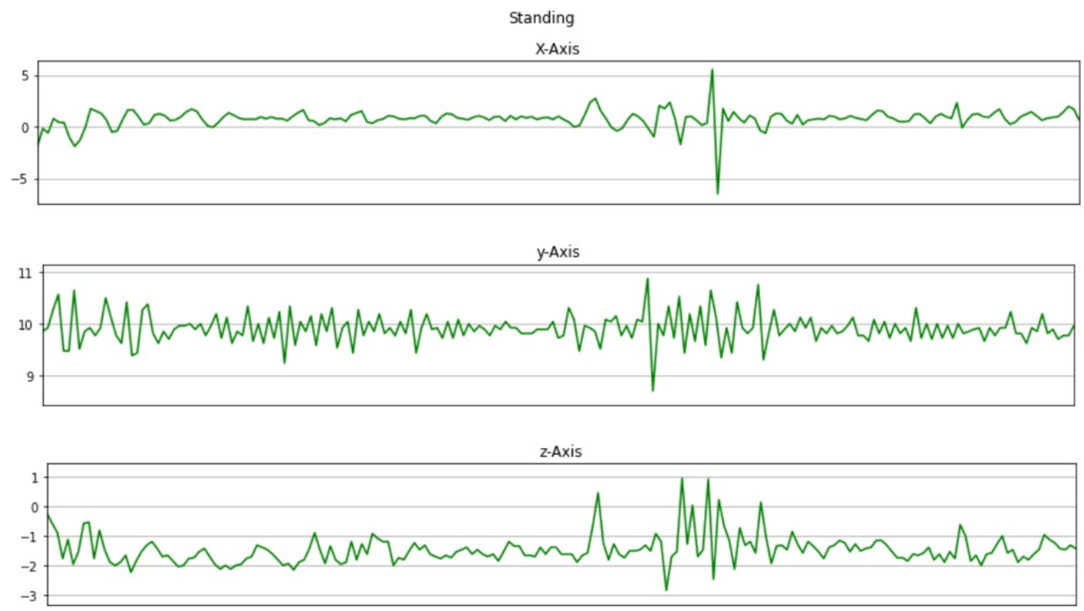


Figure 3.7: Standing category

The raw data was not balanced. There is a lot of data, the total counted data are:

- Walking (137375)
- Jogging (129392)

understanding human behavior. [40] So nowadays, the model can understand human behavior. Machine learning one of the best fields is deep learning where the model is inspired by the brain function and it is working like the brain function. [41] By the help of working like a brain function model can learn self-learning so that they can predict the behavior. A perceptron is trained by providing it with numerous training samples and computing the output for each one. The weights are modified after each sample to reduce output error, which is typically characterized as the disparity between the desired (target) and actual outputs. The perceptron's capacity to learn classification is crucial since many cognitive processes rely on classification. Similar algorithms could learn to have the same behavior as humans. So, if the model will calculate the same behaviors from the first layer it will process another second layer. The layer will pass the data consistently and finally, the output will show the result of which activity from the human.

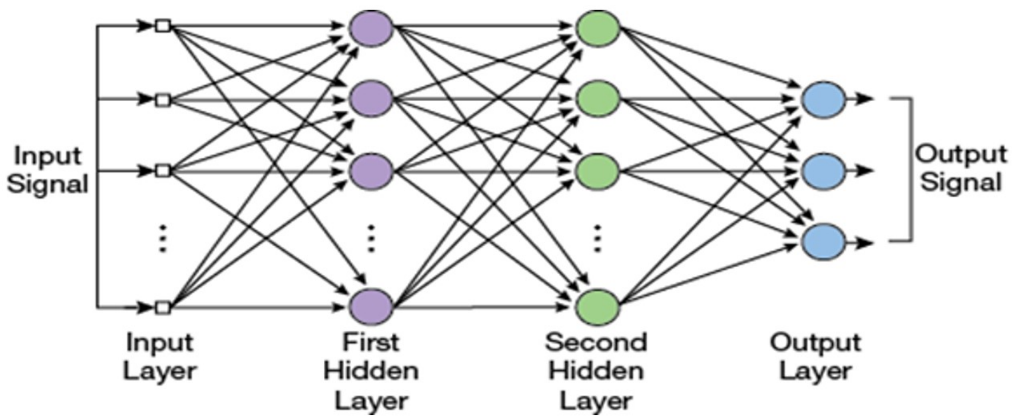


Figure 3.8: Layers from the sequential model

Sometimes there is almost similar activity from different label so in this case the layer is showing almost same but there will be different types of layer and data will go through those layers so that the input signal and output signal will be the same.

There is an arrangement sequence, so when there is a video to put into the algorithm, the frames will search for this functionality, which will be read from

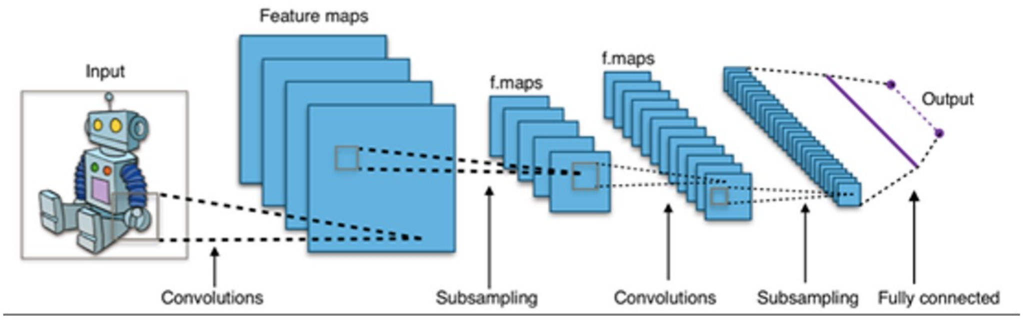


Figure 3.9: Architecture of sequential model

the videos. Then frames will be extracted from the video until the maximum frame count is reached. Most of the time there will be different frames, so when there are different areas in every video, those data will be put into the batches. For the solution of this problem, padding is used on the model. So that number of equal frames is in real shape. The image size is 224, and there are two functions on the CNN model. One will crop the center square, and another will load the video function into the CNN algorithm. Max sequence length and a number of features are important because now there will be a CNN network, which means a convolutional neural network.

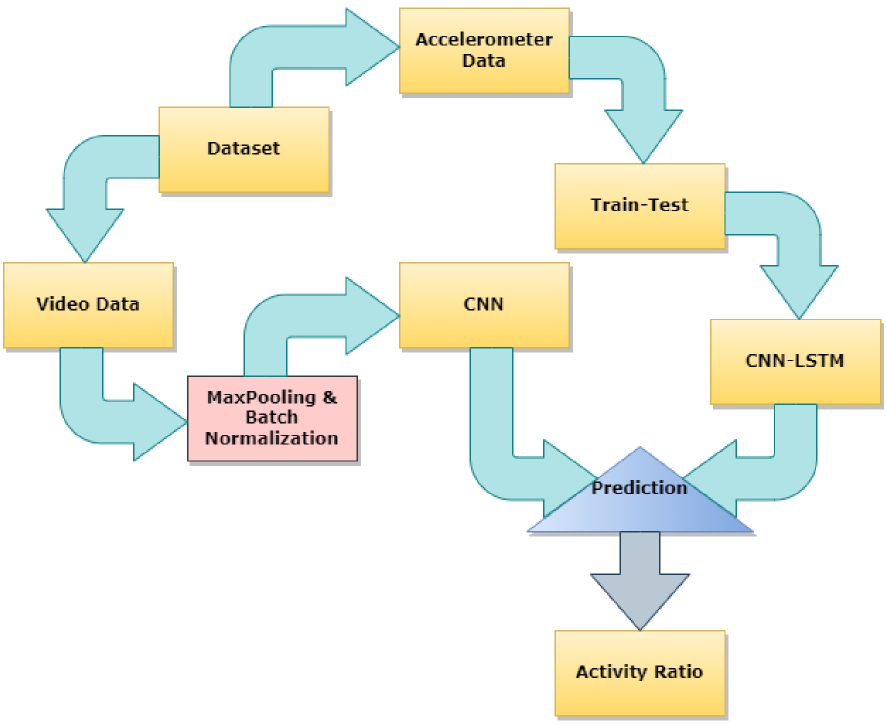
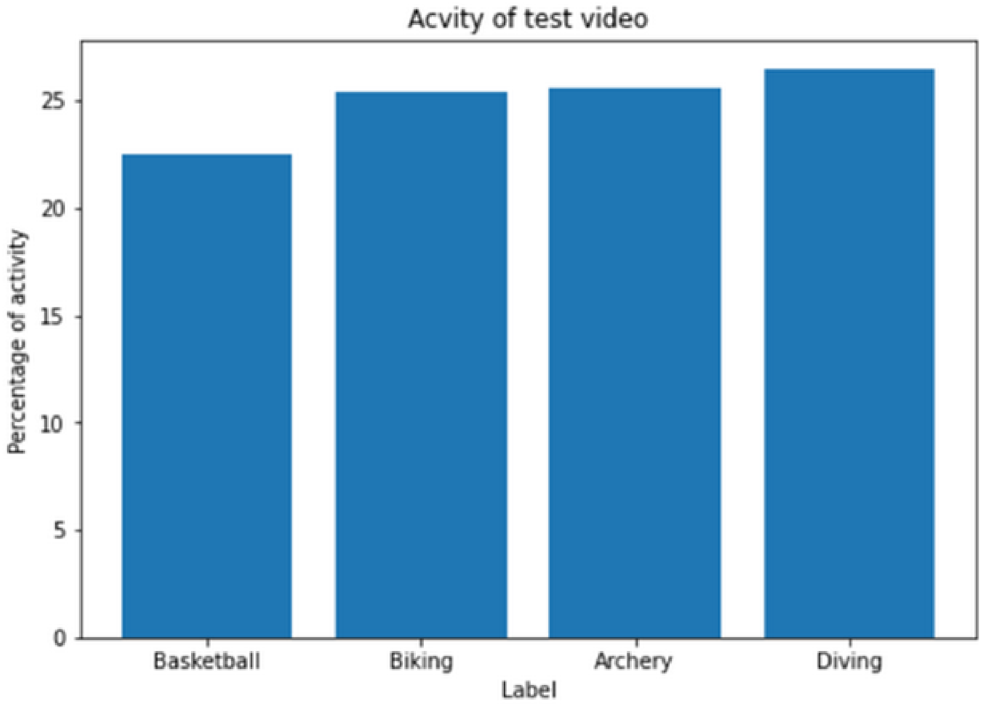


Figure 3.10: Proposed hybrid CNN-LSTM architecture to recognize human activity.



Our model's two accuracies throughout training and validation are shown in Fig. 4.7.The Keras sequential model evaluation. We measured the training and validation precision while experimenting with various epoch counts. After 10 epochs, the model achieves its maximum accuracy in training and testing as well as verification. In the epoch no. 5, the model sustains in the low accuracy and it varies all the way with no. of epoch changes. The accuracy of the model is 92%, loss=0.25.41, and value loss=0.2056.

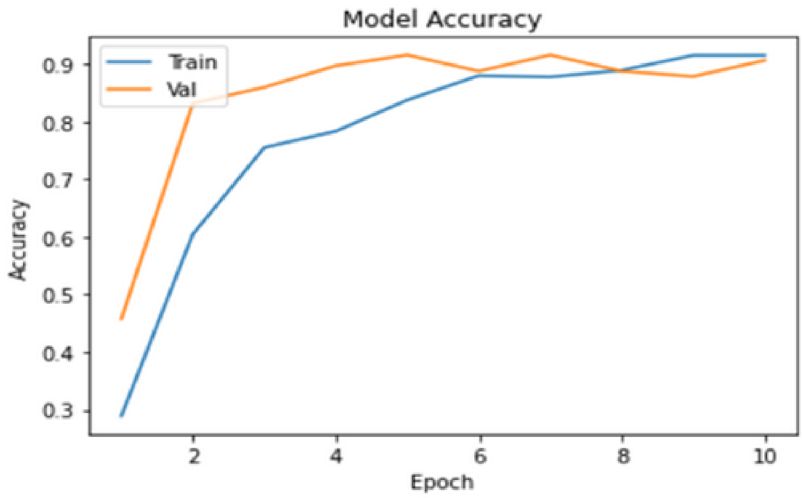


Figure 4.14: Activity of test video

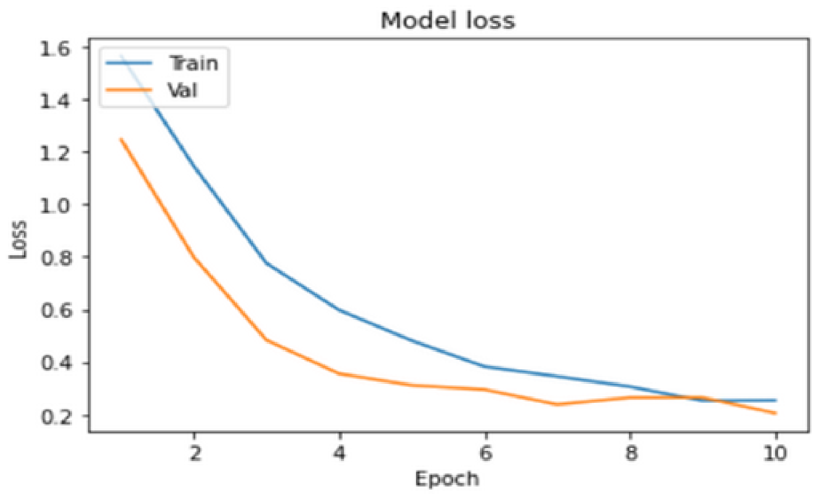


Figure 4.15: Accuracy of the Sequential model

Here is the model confusion matrix.Where we will see the output of the labeled data.

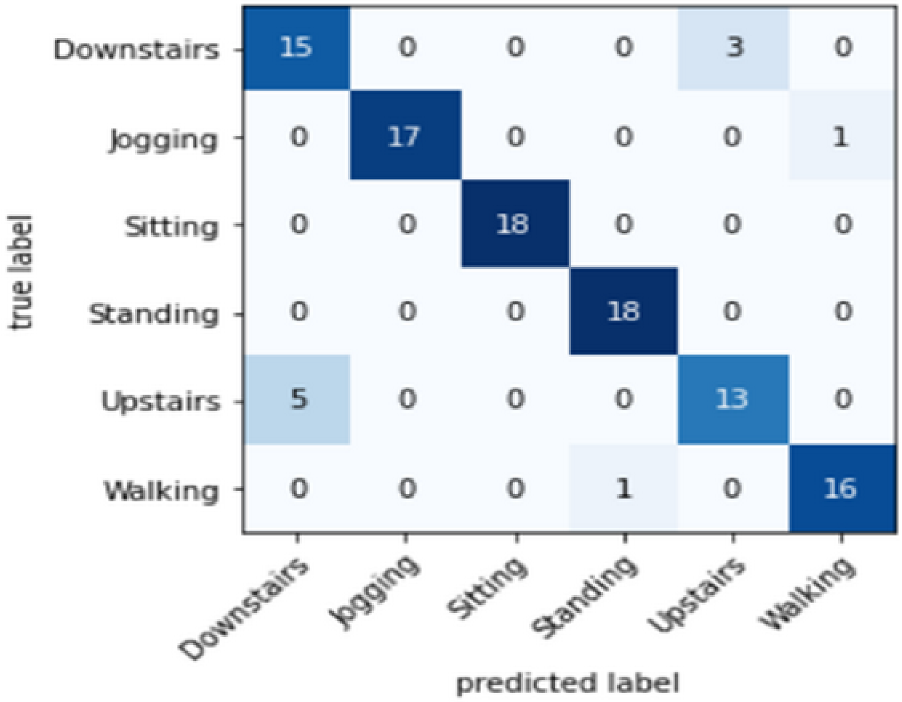


Figure 4.16: Labes of different catagory

We will now see the output of model one.Here is the showing a sample of the video:

Figure 4.17: Demo photo from video dataset

Here is the showing of the activity of the predicted label. After training the sample video then it will show the predicted output.

```
Top 5 actions:
  roller skating          : 96.85%
  playing volleyball      :  1.63%
  skateboarding           :  0.21%
  playing ice hockey      :  0.20%
  playing basketball      :  0.16%
```

Figure 4.18: Top Actions

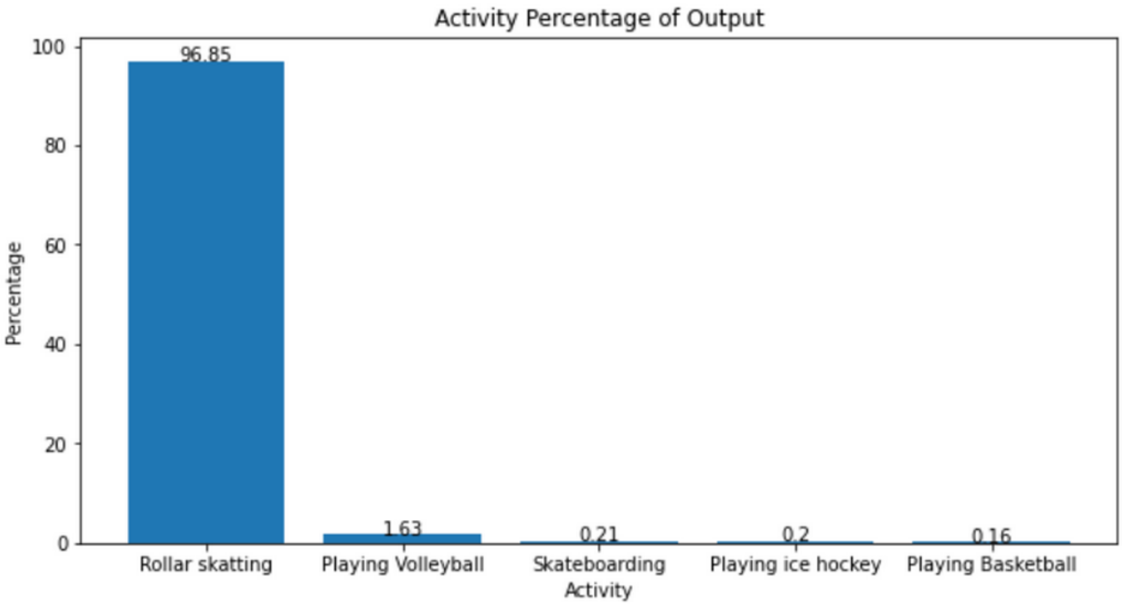


Figure 4.19: Activity Percentage of Output

Here is showing the output. Where is showing roller skating percentage is higher to predict the activity.

Appendix

6 Dataset

- UFC-101 Dataset

The UFC 101 dataset is a widely-used benchmark dataset for action recognition in videos, specifically in the context of mixed martial arts (MMA) fights. The dataset consists of 101 short video clips of MMA fights, each approximately 6-8 seconds long. The videos were collected from UFC events and feature 13 different fighters performing various moves, such as punches, kicks, and takedowns. Each video clip is labeled with one of 101 action classes, such as "punch", "kick", "takedown", "submission", and so on. The videos are captured at a frame rate of 30 frames per second, and have a resolution of 320x240 pixels. They are encoded in the MPEG-4 format and are stored as individual video files. In addition to the video data, the dataset also includes precomputed optical flow vectors, which describe the motion of pixels between consecutive frames.

Figure 6.22: UFC-101 Dataset of Basketball

Figure 6.23: UFC-101 Dataset of Biking

Figure 6.24: UFC-101 Dataset of Bowling

Figure 6.25: UFC-101 Dataset of Diving

- Accelerometer Data Set

Accelerometer datasets are commonly used in research to analyze human activity and motion. These datasets typically consist of measurements taken from accelerometers, which are sensors that detect changes in velocity or acceleration. Here are some key points that could be included in a summary of accelerometer datasets for a research paper:
Data collection: Accelerometer datasets are typically collected using wearable devices that contain one or more accelerometers. These devices can be worn on different parts of the body, such as the wrist, ankle, or waist, depending on the specific research question.
Sampling frequency: The sampling frequency of an accelerometer dataset refers to the rate at which data is collected. Common sampling frequencies for accelerometer datasets range from 20-100 Hz, although higher frequencies may be used in certain applications.
Data preprocessing: Before analysis, accelerometer datasets may require preprocessing to filter out noise or correct for sensor drift. Common preprocessing steps include low-pass filtering, high-pass filtering, and calibration. Activity recognition: One common application of accelerometer datasets is to recog-