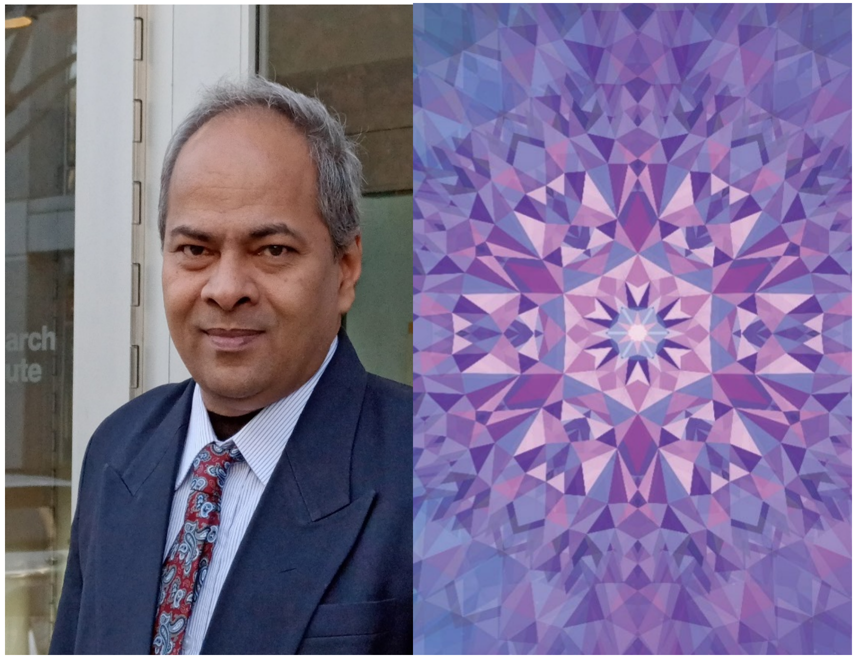
ITU KALEIDOSCOPE
ONLINE 2021

6-10 December 2021

Abnormal activity recognition using deep learning in streaming video for indoor application

Dhananjay Kumar
Department of Information Technology,
Anna University, MIT Campus, Chennai, India



Session 3: Contributions to security

Paper S3.2 - Abnormal activity recognition using deep learning in streaming video for indoor application

Outlines

- Motivation and Challenges
- Proposed Architecture
- Algorithm Development
- Experimental results
- Summary

Motivation and Challenges

Motivation:

- The growing concern about home/office safety and security
- Increase in affordability of IP-based camera systems
- The state-of-the-art system in human recognition lacks sufficient intelligence
- In streaming video, the presence of the large spatio-temporal data require intelligent approach

Challenges:

- Human activity recognition is a challenging time series classification task
- Hand-crafted features vs. deep-learning models
- The patterns of dynamics of local motions are required to be learned
- Feature selection and dimensionality reduction
- The complexity of the background, and sometimes discontinuity in the streaming video feed

Proposed System Architecture

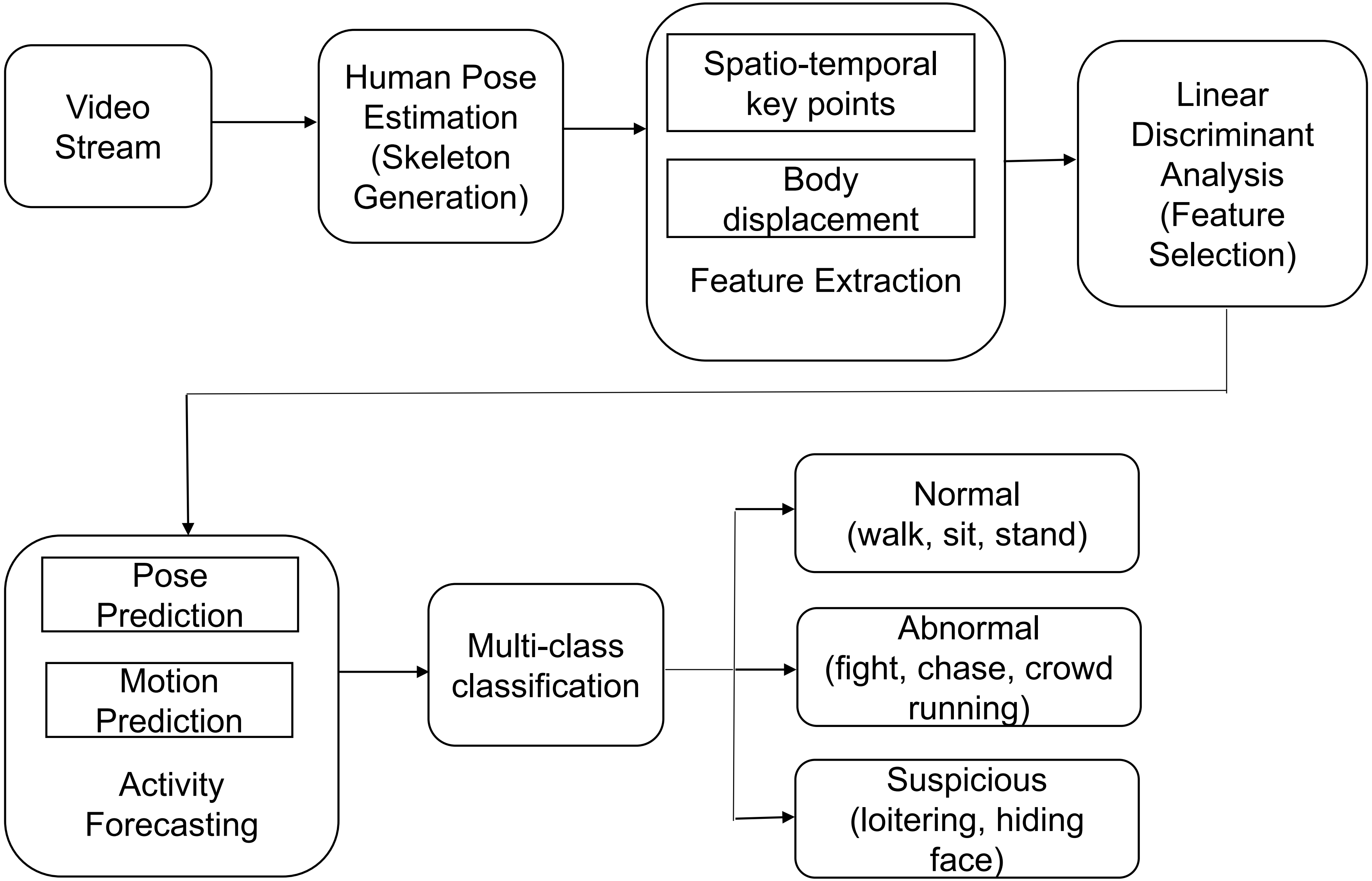


Figure 1 – System architecture

Bi-LSTM and Feature Selection

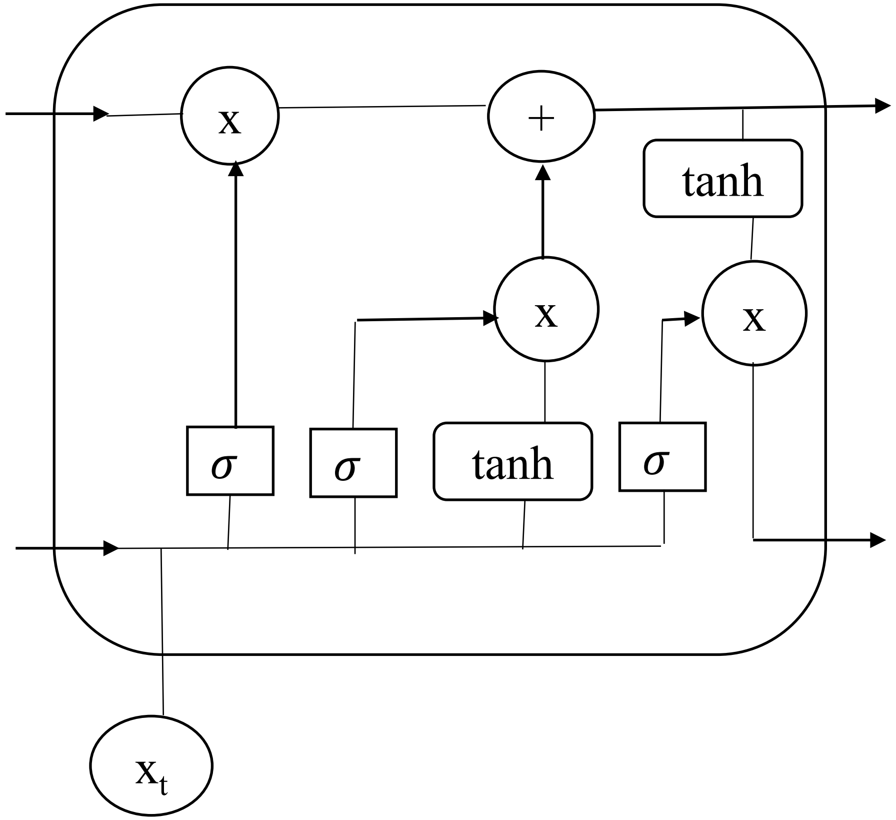


Figure 2 – A basic unit of LSTM used in Bi-LSTM

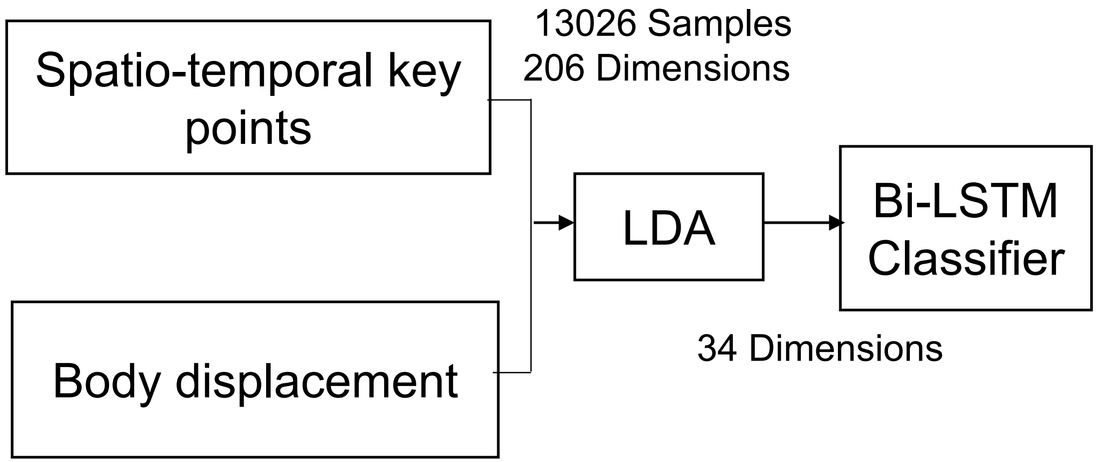


Figure 3 – Feature selection

For optimization of the features, the dimensionality reduction or feature selection methods are applied and reduced into 34 dimensions before classification

The two Long Short-Term Memory (LSTM) are stacked on top of every other, where one LSTM goes within the forward direction and another in the backward direction.

Algorithms

Algorithm 1: *Skeleton_Generation*
Input: Video stream
Initialize numbering for joints
Declare *pose_pairs*
for each frame in video:
for i in range (len(BODY_PARTS)):
Generate heatmap
Find *x,y coordinates*
if multiple person detected
Compute centroid value
for pair in pose_pairs:
Draw ellipse for *pose_pair* coordinates
Draw lines between *pose_pair* coordinates

- A continuous sequence of skeleton information is generated for a video stream
- Multiple action classes are classified by using SAF + Bi-LSTM

Algorithm 2: *SAF+BiLSTM_Train*
for X, Y in the training dataset:
n_components = min(*num_features_from_lda=Dn*, X.shape[1])
lda = LDA(n_components = n_components, whiten=True)
lda.fit(X)
X_new = self.lda.transform(X)
clf.fit(X_new, Y)
Initialize train_data with the X_new
for each skeleton sequence in X_new:
Append pose label to the data
if video pauses due to delay
Compute the current time dependent variables
$y_1(t) = a_1 + w_1^{11} * y_1(t-1) + w_1^{12} * y_2(t-1) + e_1(t-1)$
$y_2(t) = a_2 + w_2^{21} * y_1(t-1) + w_2^{21} * y_2(t-1) + e_2(t-1)$
Create SAF+BiLSTM model, Initialize the classifier
clf = BiLSTMClassifier (batch_size, timestamp, features)
Do the following until model converges:
for every pose_sequence in train_data:
predicted_score = model (*sequence_list*)
Use mean square error function to compute loss in *predicted_score*
Perform gradient descent through backpropagation
Update model weights and biases
return model

Implementation

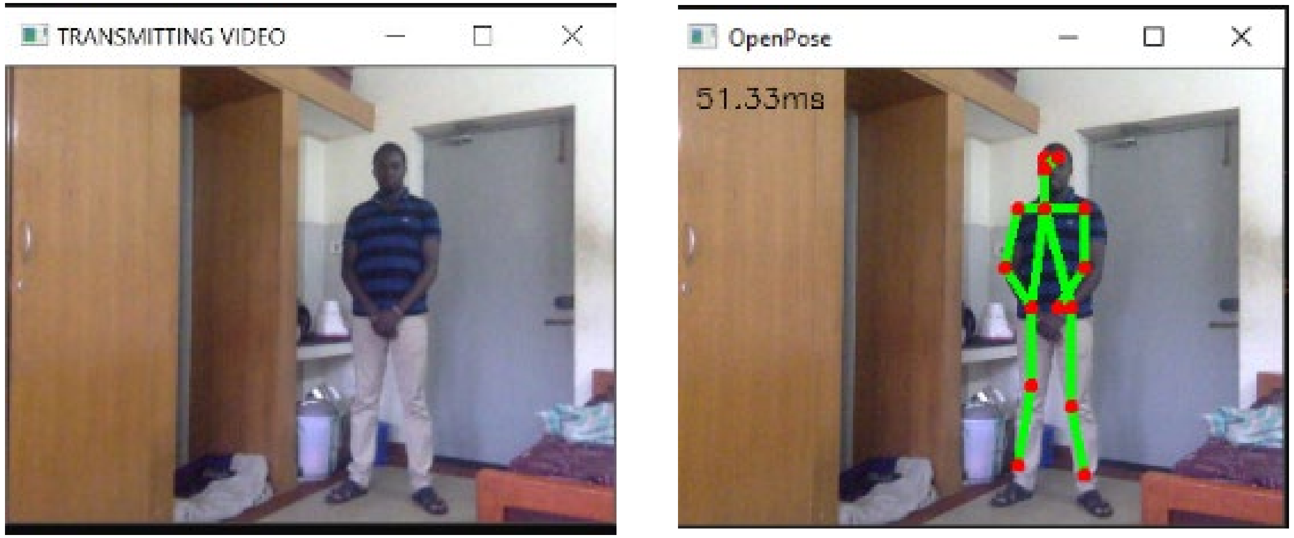
Dataset Used:

- The MSR Action Recognition
- MPII Human Pose
- IIT-B Corridor

Software/tools used

- The *OpenCV* library in *Python:* to capture and process the video stream.
- *Tensorflow* backend in *Python* using *Keras* library: to design SAF+Bi-LSTM
- The *Pickle* module in *Python*: to serialize the trained learning model

Results



a) Transmitting video **b) Receiving video**

Figure 4 – Skeleton generation

- The human skeleton is generated by using human pose estimation which is trained on the *MPII* data set
- Vector Auto-Regression (VAR), which is a multivariate forecasting algorithm, is used to perform activity forecasting
- The SAF+Bi-LSTM model classifies the human activity

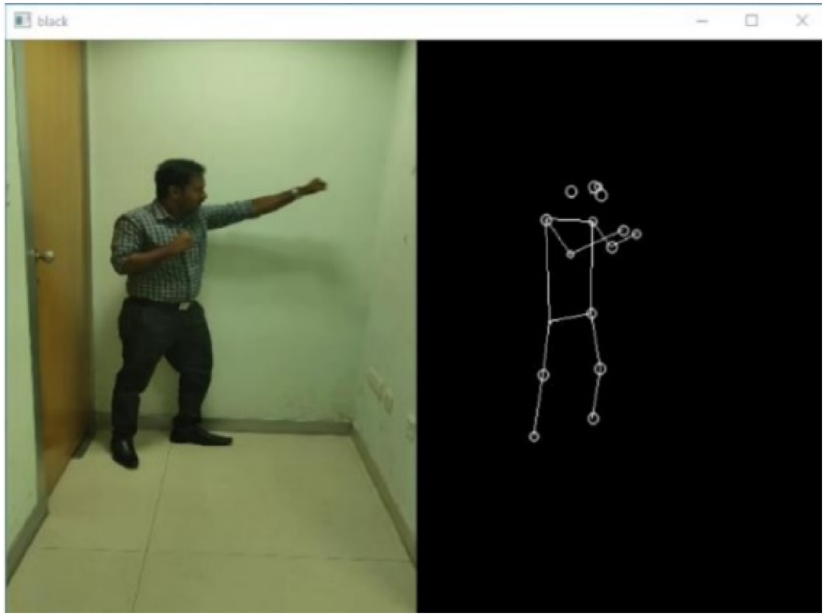
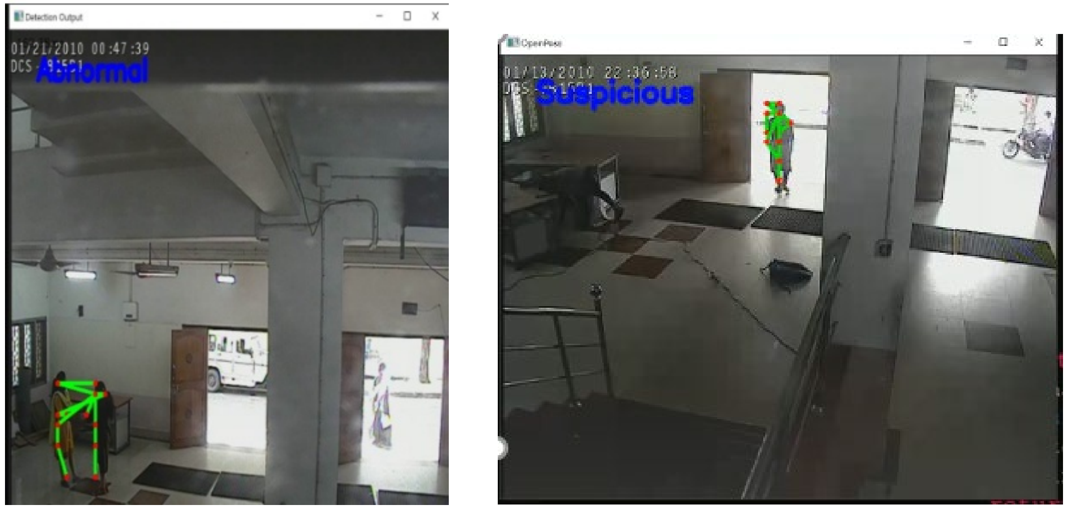


Figure 5 – Future pose prediction



a) Abnormal **b) Suspicious**

Figure 6 – Activity classification

Results

Class	Precision	Recall	F1-score	Support
Normal	0.83	0.89	0.86	5519
Abnormal	0.87	0.91	0.89	2361
Suspicious	0.87	0.9	0.88	1306

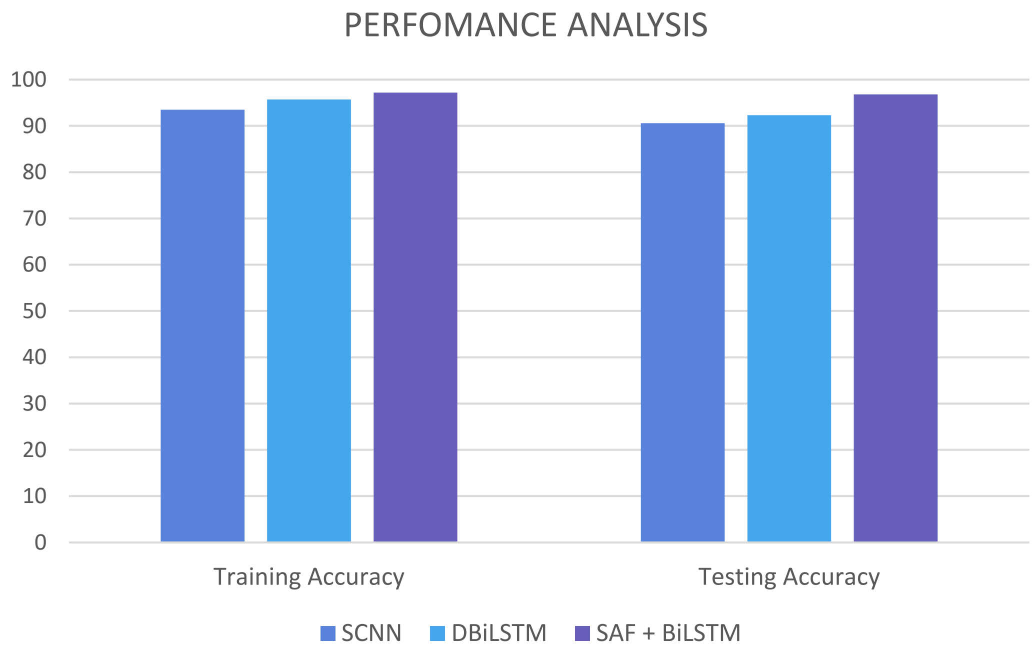
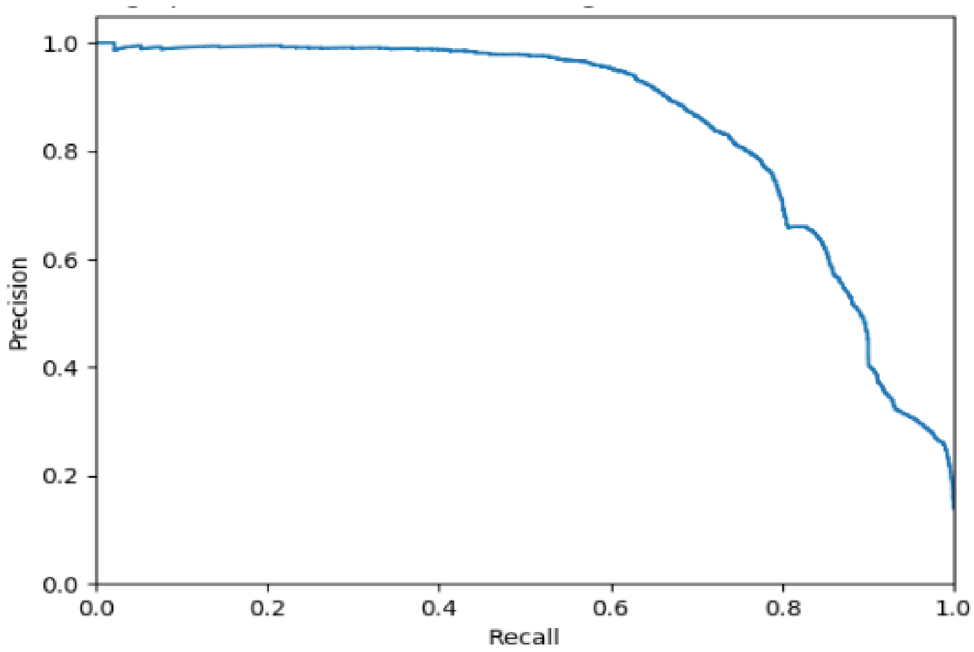


Figure 8 – Training and test accuracies



S. No.	Method	Accuracy %
1	Sequential CNN	93.5
2	DBiLSTM	95.7
3	SAF + Bi-LSTM	97.2

Summary

Conclusion:

- A deep learning-based system to identify abnormal human activities
- Combination of Skeleton Activity Forecasting (SAF) and a Bi-LSTM network
- Skeleton activity forecasting for predicting the future pose
- System performance evaluated on standard datasets

Future work:

- Further system optimization needed to support the IP based live streaming video
- It can be standardized under Recommendation ITU-T H.627 “Signalling and protocols for a video surveillance system”

ITU KALEIDOSCOPE
ONLINE 2021

Thank you!