

Human Activity Identification and Recognition Using Sequential Convolution Neural Network in Deep Learning

Mr. Kazi Azizuddin¹, Dr Premal Patel*², Dr. Chintan Shah³

Department of Computer Engineering, College of Technology, Silver Oak University,
Ahmedabad, Gujarat^{1,2,3}.
arkazi.ce@gmail.com¹,premalpatel.ce@socet.edu.in²,chintan.shah84@gmail.com³

***Corresponding author: Dr Premal Patel*²**

Email: premalpatel.ce@socet.edu.in

Abstract. Progress in human activity recognition (HAR) in video surveillance frameworks is becoming increasingly robust. There are bunches of testing confronting the video perceptions in human movement. Presently day's individuals need ideal picture of human. It is exceptionally grim task for human to perceive a movement or picture while action is performed. For this in this paper exhibiting scarcely any strategies for profound learning and highlights of human action that shows you standing out methodology or results from particular techniques. HAR can perform various exercises right now can recognize which strategy is better for HAR. Profound learning covers the procedure for neural framework which we can say it is a fake cerebrum. Neural framework gives one new opportunity to significant learning. It can perform different errands together. Arranging the physical exercises performed by a client dependent on accelerometer and spinner sensor information gathered by a cell phone in the client's pocket. The exercises to be grouped are: Handwaving, Walking, Running, Boxing, Hand clapping etc.

Index Terms- Deep learning, Neural Network, CNN, Activity, and HAR.

I. INTRODUCTION

Deep learning covers the technique for neural system which we can say it is a counterfeit cerebrum. Neural system gives one new chance to profound learning. It can perform various tasks together. It has three fundamental sorts (1) Feed forward neural system (2) Recurrent neural system (3) convolution neural system. It is a one kind of structure or complex system ornamentation that plays out all activity on neurons. Neurons are lies in cerebrum. It gives huge measure of security information. Furthermore, that security information is successively in way of high processor. It has three kind of layer: input layer, output layer, hidden layer. Each neuron has them possess information by utilizing their standard or utilizing their calculation which human arranged to it. It is a genuine program that created by claim self. Significantly it performs six-seven calculation techniques. By including those strategies heaps of various applications are in showcase. Also, HAR is one from that. HAR is a wide Application of Deep learning and playing out the activity of picture preparing. Fundamentally profound learning has three learning techniques: (1) administered realizing which performs naming information (2) unaided realizing which has without marking perform (3) semi directed which is less data is known n less is obscure. Utilizing profound adapting at present research on reconnaissance video framework, wise vehicle framework, remote detecting, Moving article following, Defense observation framework. They give PC communication human and give individuals best outcome to play out investigate the errand of human movement. Profound learning is a dependent on neural system techniques. Neural system is

amazing arrangement of profound learning. Those techniques can improve human-PC connection. Each activity performs by people and mix of activity work in movement. Perceive singular activity perform action. So it is a huge theme of subclass variety of HAR.



Figure 1. Hand Waving Sample vide

II. Literature Review

In this paper author [1] trying to say that recognize multiple people in one activity it is difficult task to recognize and larger subtopic of image processing. So, this remains very large intensity challenging topic of intra-class variation of human activity. For this problem solve multiple people tracing they scrutiny on information about context. Context is defined as information that is not directly related to human. They are using spatio-temporal volume descriptor to capture nearby person actions. And other is deep models network to capture the contexts for HAR. Deep neural network involving chance of variation in reasoning power and constitute multiple sources of context information. Therefore, they proposed a deep neural network based on recognize HA. They introduce 2-level scene context descriptor. That describes territory information of centered target at the world wide and in particular area of one's level. They jointly seize the multiple sources of context information.

In this paper author Mohanad Babiker [2] proposed to work recognize human's daily activity by video surveillance system. In this paper they mention that this system is capable to monitor every moment without blowing down the eyes. For that they are using concept of image processing in background subtraction, binarization, and morphological operation. They describe video surveillance system performed by human semi, automated and fully automated. They perform preprocessing operation to image for making more intelligible to computer. They use first static frame. For that they are using multilayer preceptor feeds forward neural network. Neural network for train the design system for connectivity of input and numeric relation with specific output through a number of neurons nodes. Their goal is to measure the accuracy of the designed system by observing the validity of the system to recognize human activity. They want to design more complex activity which has high similarity in human shape and movement because on this paper they recognize 5 activity which are walking, lying, hand waving, boxing and sitting.

In this paper author Sarita Chaudhary, Mohd Aamir Khan, Charul Bhatnagar [3] focused on main 3 things moving object detection, object tracking, behaviors understanding for activity recognize. They focus on speed, direction, centered and dimensions for feature extraction for tracking objects. In this paper author describe only anomalous or normal activity. They design a system to handle the abnormal conduct in video. They used Rule based approach for classifying events as normal and abnormal. Motion pattern are examine and nearly new as features like they follow commanding work. Ascendant activities in video they cover

abnormal activities but they work like multiple events like tossing a coin twice. They have headed first time and tail at second time. Discovering exercises agreeing acknowledgment. They give labeling to all events. They perform running, walking and crawling. Application of anomalous event detection system is: traffic monitoring, medical science, highly surveillance areas like military area, airport etc., crowd analysis, criminal activity recognition, automatic forensic video retrieval. GMM (Gaussian mixture model) is used for object detection and some problem domain rules are used distinguish different types of behavior. Rule-based method rules are established, and that rules helps to categories the sample as normal or anomalous. Representation of normal regions in video stream is quite difficult. Exact notion of abnormal behavior is either subjective or changes over different applications. Probability of vague limit between typical or strange conduct. Nearness of clamor and dubious to name the conduct. Article conducts seeing exceptionally testing assignment. Since object conduct can't be predicated straightforwardly from an information video. For future they want to perform (1) they want to focus on some more suspicious activities (2) multiple objects Overlapping will be taken in consideration.

In this paper writer [4] wants to make one radar system which is recognize limb motion by SGRUN technique. And then they are checking with LSTM and did comparison of GRU, SGRUN and LSTM. In this radar system they are work under electromagnetic waves they unconstrained by the lighting condition and raise no privacy concern. As they are said that they perform on electromagnetic waves in that frequencies are there for recognize motion of body. Frequencies are in two form Doppler and micro-doppler.in paper they mentioned Doppler frequencies are changed by time but micro-doppler changed with motion of the other body parts as limb. For that they [4] are using STFT (short time Fourier transform) frequency analysis method. In this paper they mentioned for sequential learn data so for that they try different methods of deep learning they seen some variation of RNN then LSTM then GRU. They proposed new radar system SGRUN for classify motion. RNN is good for sequential data compare to CNN. It is preferred for choice to learn the time-varying motion signatures and to improve the human body and limb motion classification. GRU is simultaneously capable to remember existing feature in the input stream. (1) They perform different methods in that they mentioned that LSTM and GRU are weak in gradient vanishing problem. In LSTM there is more parameter and hard to train data. But in GRU is fewer parameter and easy to train. In the future author collect a more complete dataset with more subjects and more motion situations to do more work on the human body and limb motion recognition exploiting radar.

In this paper author Jian sun, yongling Fu, Shengguang Li, Cheng Xo and Lin Tan [5] want to propose a hybrid deep framework based on convolution operation, LSTM recurrent units and ELM classifier for sequential recognize human activity. In this paper they have study on different types of sensor such as audio sensor, image sensor, accelometer. In this paper they are study HAR based on wearable sensors. Mainly they are focus on improve recognition accuracy and increase complex accuracy. In this paper they are search every possible method for recognition and then go with deep convolution network and LSTM.LSTM which is better to solve in sequential data so it gives better result but complex activity is suffer so for that they integrate extreme learning machine in too deep convolutional network. So basically they perform CNN, LSTM, and ELM which can perform sequential activity. In this paper they want to show that deep framework is using with ELM classifier is best compare to fully connected layer. CNN are able to learn hierarchical data representation on classifier that leads to perfect analysis of data. LSTM is the extension of RNN that uses memory cell for long term.[5] Here they are used LSTM and extracted feature by CNN in particular time series. Feature extraction automatically performs without any intervention. In CNN they are said and facing problem of time series problem. It is good for feature extraction and classification.

Sharing units in CNN between multiple sensors.[5] OPPORTUNITY benchmark dataset they are using. They want to perform transfer learning approach based on existing models recognizes activity and they want to perform this on large-scale data.

In this paper author [6] proposed to recognize both space and time information by RGB input data. They are performing this by MSTRNN method and for more convolutions they are performing fully connected layer. They perform Batch normalization.[6] First they are taking 3 human datasets and compare them with those methods check capabilities and accuracy. Their aim is to observe internal representation of properties. MSTRNN receive a stream of RGB images in the input layer. Training layer depends on the output of the previous layer, but the distribution of neuronal activations generated by every layer changes after each update of the weights. This causes training of deep neural network to slow and difficult. For that they are using CL1AD, CL2AD, 3ACWD dataset. Author wants to perform more accurate and effective model. They want to perform more areas of images that are not relevant to given task. Here they want to perform on larger dataset. In this paper author Mohammed Mehedi Hassan, Md. Zia Uddin, Amr Mohamed and Ahmad Almogren compared with traditional method to artificial neural network methods like SVM and ANN. In this paper they present smart phone sensor based approach for recognize human activity. They recognize normal activity which are walking, reading, standing, lie-to sit, sit-to-lie etc. they define activity categorized in to 3 groups short activity, simple activity and complex activity.[7] Here they are performing basic activity which is walking and reading. Here in training over fitting problem is occurring to make it robust they are using KPCA and LDA (linear discriminate analysis). They recognize activity on wearable sensors they recognize every present activity and require no static infrastructure to operate it. They are using DBN for activity recognition it is good to recognize human activity. For future author want to plan for more robust features and learning for more efficient and complex activity's recognition in real time environment. In this review paper author define various problems of classification and evaluation process of human activity recognition. They define various deep learning methods and their challenging task. [8] They are suggesting over fitting and under fitting problem for PCA, LDA, and ECDF. for large activities they define Support vector machine, hidden markov model, decision tree ,k-nearest tree, those are traditional methods used for background subtraction. Then they are suggesting for motion sensors gyroscope, magnetometer, and accelerometer. Then after they are define current trends for deep learning methods of RNN, restricted Boltzmann machine, sparse coding etc. They are giving introduction of deep learning methods. In this paper authors are shown deep learning methods have their advantage [8]. They are talking about which methods is best, which dataset and then which software we have to perform those methods and why they are reliable. They have ability of deep learning methods that they are automatically learn from unlabelled raw sensor data. They are using different types of dataset. Then they are shown different databases for different types of activities. BENCHMARK dataset when we want to study on previous activity or compare to previous activities. OPERTUNITY dataset is used for daily living activities, like routine activities, and then THE SKODA MINI CHECKPOINT DATASET is for gesture. One person can hold different types of gestures. DAILY AND SPORT ACTIVITY this is for classification using on body by sensors placed on different parts of the body. There are many other datasets which is WISDM dataset, PAMAP2, MHEALTH. For future author want to proposed more deep learning techniques and machine learning techniques gives more implementation on wearable mobile sensors methods for human activity recognition.

In this paper author Chandrashekar M Patil, Jagadeesh B and Meghana M N are detect and track of moving objects and activity recognition of object in video surveillance system [9]. In

this paper they used background subtraction for Gaussian mixture model and Graph cut method for large amount of calculation is involved. Then they make a frame of video. After they are try to perform recognize activity by these model HMM, SVM, CRF. they are used HOG for extract feature from an image .For future author want to reduce false detection like shadow other reflections human, it can also incorporate HAR for moving background scenario and recognizing various activity like talking, eating etc. they are using UT-interaction dataset.

In this paper authors are discuss about more wearable sensors with methods of machine learning and data mining they said that the approach of machine learning and data mining combine are giving more reliable solutions. By machine learning they solved many challenging problem of human activity recognition.[10] Sreenivasan Ramasamy Ramamurthy and Nirmalya Roy discuss about activities which are functional activities and ambulation activities. First they define traditional approach of ML technique like SVM; HMM they are used in AR. Then after as environment become complex they introduce for deep learning because they give heuristic result of problem. Deep learning learn feature directly from the data. Then after they give lists of datasets which are used in different types of activities. So basically solve problem of human activity recognition they give different approaches i.e. transfer learning in AR, active learning in AR, deep learning in AR, semantics in AR. For future they said about various filed of deep learning and said activity learning by transfer learning is new generation of HAR.

In this paper author proposed method of HAR from raw accelerometer with RNN and they combine various architecture parameters for best parameter value. They said RNN is suitable for time series sequential data like audio and video. They compared RNN with different traditional methods for fast execution. Then after increase accuracy of RNN they performed various parameters. Then investigate the factors which are affecting the accuracy [11]. Then they compare CNN and RNN in that they said CNN is not suitable for real time execution. Then Masaya Inoue, Sozo Inoue and Takeshi Nishida are discussing about throughput. They want to train raw accelerometer high throughput data with the RNN. They are using LSTM for reduce error rate. For giving input and output there is problem occur of vanishing gradient problem so they are uses CEC (constant error carousel) it is solved by this method. Then they are checked for DRNN it is DEEPRNN for high throughput activity recognition. They constructed 3 axis accelerometer for time duration and each has LSTM unit.RNN can be suitable for time series data. RNN has high throughput that can deal with raw sensors data with fast execution. They are using HASC (human activity sensing consortium) this data set is also divided in two types of data sequential data, segmented data. For additional dataset they take HAR dataset [11]. For future author said for more accuracy, techniques such as forget mechanism, and pre-training, optimization and other methods which perform sequential data they want to studied that.

In this paper Wenchao XU, Yuxin Pang, Yanqin Yang and Yanbo Liu [12] are constructing CNN with raw accelerometer. They are constructing CNN for human activity Recognition. And then they are compared this method with SVM and checked accuracy. Authors want to achieve high recognition with low cost. They are check multiple daily activities with datasets and recorded activities with 4 types of different sensors.SVM based methods can be calculated easily. And overall accuracy count by CNN confusion matrix [12]. Authors want to perform feature extraction on SVM with high computational cost. And compared to that CNN achieved with low computational cost.[12] For future authors want to more research and improve their performance on complex detection .for that they are using WISDM dataset.

III. Overview of Different Method

3.1 Neural Network

Neural Network systems are utilized for tackling numerous business issues, for example, deals estimating, client inquire about, information approval, and hazard the executives. For instance, at Statsbot we apply neural systems for time-arrangement expectations, irregularity discovery in information, and normal language understanding.

3.2 Convolution Neural Network

Neural systems offer various favorable circumstances, including requiring less formal factual preparing, capacity to certainly recognize complex nonlinear connections among reliant and free factors, capacity to identify every single imaginable collaboration between indicator factors, and the accessibility of different preparing layers.

3.3 Recurrent Neural Network

Intermittent neural systems, otherwise called Recurrent Neural Networks, are a class of neural systems that permit past yields to be utilized as data sources while having concealed states. usually known as Recurrent Neural Network is an exceptionally well known Deep Learning model which is utilized to do various Deep Learning assignments like Time Series expectation, Image Captioning, Google auto complete component, and so forth. RNN as the name proposes utilizes recursion procedure to assemble models.

3.4 Deep Belief Network

In profound learning, a profound conviction arrange is a generative graphical model, or on the other hand a class of profound neural system, made out of various layers of idle factors , with associations between the layers yet not between units inside each layer. At the point when prepared on a lot of models without supervision, it can figure out how to probabilistically recreate its data sources. The layers at that point go about as highlight identifiers. After this learning step, a DBN can be additionally prepared with supervision to perform characterization

3.5 Convolutional Sparse Coding

Convolution Sparse Coding is a technique for learning to capable of production or reproduction models as translational invariant word references for an enormous wide range of preparing signals. Those related models have been demonstrated viable for taking care of issues in neural and mind data preparing. Convolution meager coding, which replaces the straight blend of a lot of word reference iotas with the entirety of small arrangement of mapping term tangled with the lexicon channels, was proposed to improve the application execution of conventional scanty coding.

IV. Proposed Architecture

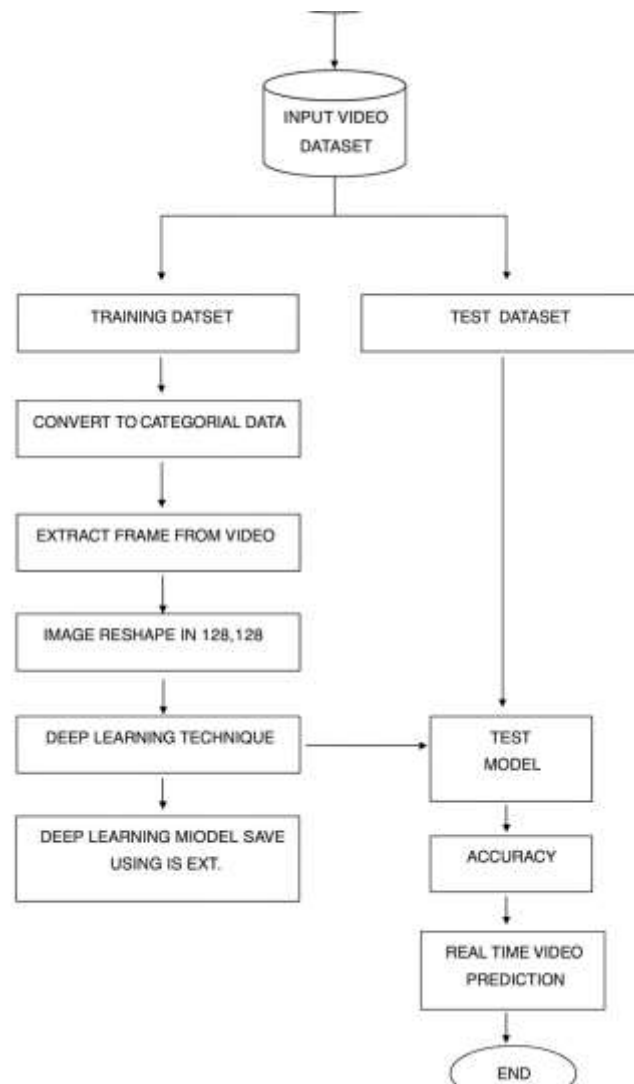


Figure 2. Proposed Architecture

```

from tensorflow import keras
model=keras.Sequential()
model.add(keras.layers.InputLayer(input_shape=(128,128,1)))
model.add(keras.layers.BatchNormalization())
model.add(keras.layers.Conv2D(16,kernel_size=(3,3),activation='relu'))
model.add(keras.layers.MaxPool2D(pool_size=(2,2)))
model.add(keras.layers.Dropout(0.5))
model.add(keras.layers.Dense(5,activation='softmax'))
model.compile(optimizer=keras.optimizers.Adam(learning_rate=0.001),loss
='binary_crossentropy',metrics=['accuracy'])
model.summary()

```

```

50/50 [=====] - 29s 574ms/step - loss: 0.1921
- accuracy: 0.7831
Final Accuracy is : 78 %

```

CNN Model-1

```

from tensorflow import keras
model=keras.Sequential()
model.add(keras.layers.InputLayer(input_shape=(128,128,1)))
model.add(keras.layers.BatchNormalization())
model.add(keras.layers.Conv2D(32,kernel_size=(3,3),activation='relu'))
model.add(keras.layers.Dropout(0.5))
model.add(keras.layers.Dense(5,activation='softmax'))
model.compile(optimizer=keras.optimizers.Adam(learning_rate=0.001),loss
='binary_crossentropy',metrics=['accuracy'])
model.summary()

50/50 [=====] - 25s 504ms/step - loss: 0.1522
- accuracy: 0.8120
Final Accuracy is : 81 %

```

CNN Model-2

```

from tensorflow import keras
model=keras.Sequential()
model.add(keras.layers.InputLayer(input_shape=(128,128,1)))
model.add(keras.layers.BatchNormalization())
model.add(keras.layers.Conv2D(64,kernel_size=(3,3),activation='relu'))
model.add(keras.layers.Dropout(0.5))
model.add(keras.layers.Dense(5,activation='softmax'))
model.compile(optimizer=keras.optimizers.Adam(learning_rate=0.001),loss
='binary_crossentropy',metrics=['accuracy'])
model.summary()

50/50 [=====] - 27s 560ms/step - loss: 0.1412
- accuracy: 0.8473
Final Accuracy is : 84 %

```

Proposed CNN Model

V Results



Figure 4. Input Video (Handwaving)



Figure 5. Input Video (Running)



Figure 6. Labelled Output Video
(Running)



Figure 7. Labelled Output Video
(Handwaving)

VI. Comparison Chart and Table

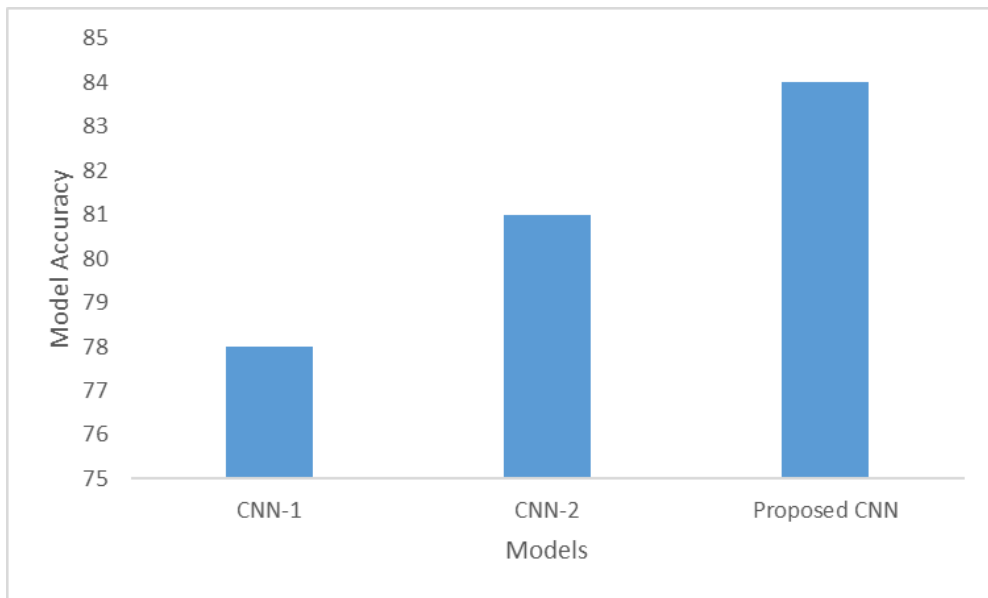


Figure 8. Comparison Chart

Algorithm	Accuracy
CNN-1	78
CNN-2	81
Proposed CNN	84

Figure 8. Comparison Table

VII. Conclusion

In the present situation by as indicated by this framework we are attempted to make separate various sorts of system layers and system which are under the zone of profound learning. We distinguished various kinds of intermittent layer and their proposed work in various situations. We proposed to make Human action and perceive two-sorts of action which are associated with one another. We attempt to perceive and recognize those items which are in picture with video and given better outcome with precision. In this manner here we infer that CNN we are utilizing with various time and various layers with their accessibility with information. So we are tried to cause one to propose framework with this CNN to perceive human activity. And I got 84% of accuracy for identifying HAR.

REFERENCES:

- [1] Li Wei and Shishir K. Shah, "Human Activity Recognition using Deep Neural Network with Contextual Information", VISIGRAPP, 2017
- [2] Mohanad Babiker, Othman o. khalifa, Kyaw Kyaw Htike, Aisha Hasssan, Muhamed Zaharadeen, "Automated Daily Human Activity Recognition for Video surveillance Using Neural Network", ICSIMA, 2017.
- [3] Sarita Chaudhary, Mohd Aamir Khan, Charul Bhatnagar, "Multiple Anomalous Activity Detection in Videos", ScienceDirect, ICSCC, 2018.
- [4] Mingyang Wang, Guolong Cui, Xiaobo Yang, Lingjiang Kong, "Human body and limb motion recognition via Stacked gated recurrent units network", IET Journal, 2018.
- [5] Jian sun, yongling Fu, Shengguang Li, Cheng Xo, Lin Tan, "Sequential Human Activity Recognition Based on Deep Convolutional Network and Extreme Learning Machine Using Wearable Sensors", HINDAWI journal, 2018.
- [6] Haanvid Lee, Member, IEEE, Minju Jung, and Jun Tani, "Recognition Visually Perceived Compositional Human Actions by Multiple Spatio-Temporal scales Recurrent Neural Networks", IEEE, 2017.
- [7] Mohammed Mehedi Hassan, Md. Zia Uddin, Amr Mohamed, Ahmad Almogren, "A Robust Human Activity Recognition System Using Smartphone Sensors and Deep learning", ELSEVIER, 2017.
- [8] Henry Friday Nweke, Ying Wah Teh, Mohammed Ali Al-garadi, Uzoma Rita Alo, "Deep Learning Algorithms for human activity Recognition using mobile and wearable sensor networks: State of the art and research, ELSEVIER, 2018.
- [9] Chandrashekar M Patil, Jagadeesh B, Meghana M N, "An Approach of Understanding Human Activity Recognition and Detection for Video Surveillance using HOG Descriptor and SVM Classifier", IEEE, 2017.
- [10] Sreenivasan Ramasamy Ramamurthy, Nirmalya Roy, "Recent Trends in machine learning for human activity recognition- A survey", Department of information system, University of Maryland, WILEY, 2018.
- [11] Masaya Inoue, Sozo Inoue, Takeshi Nishida, "Deep recurrent neural network for mobile human activity Recognition with high throughput", Springer, 2017.
- [12] Wenchao XU, Yuxin Pang, Yanqin Yang, Yanbo Liu, "Human Activity Recognition Based On Convolutional Neural Network", IEEE, 2018.
- [13] https://www.google.com/search?q=anomaly+detection+of+human&rlz=1C1CHBF_enIN824IN825&sxsrf=ACYBGNQOS6EFyYPIgl1C6pA7DYZPPvjhSA:1570444237381&source=lnms&tbn=isch&sa=X&ved=0ahUKEwiyiKLX-InIAhWnVt8KHSxYCQQQ_AUIEigB&biw=1365&bih=598
- [14] Nils Y. Hammerla, Shane Halloran, Thomas Pfotz, "Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables", IEEE, 2018.
- [15] Wilton WT Fok, Louis CW Chan, Carol Chan "Artificial Intelligence for Sport Actions and Performance Analysis using Recurrent Neural Network (RNN) with Long Short-Term

Memory (LSTM)", IEEE, 2018.

- [16] Yusen Wang, Wenlong Liao, and Yuqing Chang, "Gated Recurrent Unit Network-Based Short-Term Photovoltaic Forecasting", energies, 2018.
- [17] Xin Wang, "OGRU: An Optimized Gated Recurrent Unit Neural Network", IOP, 2019.