Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

Pose based Activity Recognition using Supervised Machine Learning Algorithms

[1]Siddharth Tomar, [2]Anmol Kumar Sharma, [3]Tina, [4]Kapil Gupta
[1,2,3]Student, National Institute of Technology, Kurukshetra
[4]Assistant Professor, National Institute of Technology, Kurukshetra

***Abstract*: Human Activity Recognition is turning into a well-known field of exploration over the most recent twenty years. Understanding human conduct in pictures gives helpful data for a huge number of PC vision issues and has numerous applications like scene acknowledgment and posture assessment. There are different strategies present for action acknowledgment. In this postulation, methodology for human movement acknowledgment utilizing an individual's posture skeleton in pictures is proposed. This work is separated into two sections; a single individual pose assessment and activity classification. Pose Estimation comprises the acknowledgment of 18 body key points. Then, action grouping task is performed by utilizing various calculated relapses. Additionally, correlation between different other regression and algorithms' accuracy on dataset is shown. To assess the error during the pose assessment we utilize the Euclidean distance equation which brings about a normal inexact mistake of 7.166. For the action characterization task, we utilized six different supervised machine learning algorithms in which Decision Tree and Random Forest give the most noteworthy exactness of 96% and Naive Bayes gives minimal precision of 42.66%. For dataset prerequisites, we have made our own dataset with the assistance of the Roboflow system. Arrangement of dataset is done by separating into two sections, one is utilized to prepare the model and one more is utilized to approve our proposed model's presentation. We will likewise examine a few benefits and hindrances came across project and view a concise examination of exactness. The discoveries will likewise show how the vision-based methodology is turning into a well-known methodology for HAR research nowadays.**

Index terms: Human Activity Recognition (HAR), Pose Estimation, OpenPose, Computer Vision, Supervised Machine Learning Algorithm

I. INTRODUCTION

Human Activity Recognition (HAR) can be characterized as a method of deciphering still pictures, recordings, and tactile information to order a progression of human exercises [7], [13]. HAR has turned into an amazingly renowned logical theme in the PC vision local area. It is locked in with the progression of various huge applications, for instance, human-PC association, increased reality, security, video reconnaissance, and wellbeing observing. Aside from its application, it actually viewed as a troublesome errand to achieve due to a few perplexing worries, for example, gadget or sensor movement, foundation mess, impediment, inconstancy in scale, lighting, and so on [8]. HAR framework generally relied upon administered and solo learning. In Supervised learning, we train the machine using data that is very much named. Unsupervised learning is an Artificial Intelligence (AI) procedure, where you don't need to manage the model. Taking everything into account, you really want to allow the model to manage its own to track down information. In the pose-based technique, exercises are arranged to utilize assessing the individual's body joints through Convolution Neural Network (CNN) [5], [9], [10]. In HAR one of the essential difficulties is an impediment and staying away from impediment. The objective of a HAR framework is to anticipate the name of an individual's activity from a picture or video. This intriguing subject is roused by numerous valuable genuine applications, like reproduction, visual reconnaissance, understanding human conduct, and so on. Action alludes to the development of the whole body or the various places of the appendage's comparative with time against gravity. Human movement acknowledgment assumes a significant part in the cooperation among individuals and relational connections. Human Activity Recognition (HAR) turns into an extremely well-known and dynamic examination region for analysts over the most recent twenty years. It gives data like a singular character, mental state, and character. Activity acknowledgment through recordings is a notable and set-up research issue. Conversely, picture-based activity acknowledgment 6 is a similar, less investigated issue, yet it has acquired the local area's consideration lately. Since movement exercises can't be assessed through the still picture, acknowledgment of activities from pictures stays a dreary and testing issue. It requires a lot of work as the methods that have been applied to video-based systems cannot be applicable in this. However, the approach is not the only problem faced in this task. There are many other challenges, too, especially the changes in clothing and body shape that affect the appearance of the body parts, various illumination effects, estimation of the pose is difficult if the person is not facing the camera, definition, and diversity activities, etc. As a result of this study, many applications, such as human computer interactions, video surveillance systems, and robotics that characterize human behavior, require multi-activity recognition systems [12]. Among the various techniques for grouping, two fundamental inquiries emerge: "What activity?" (Intellectual issue) and "Where it is in the picture?" (Limitation issue). When attempting to perceive human exercises, it is important to decide the condition of human development (Kinetic state) with the goal that the PC can adequately perceive these day-to-day routines and are somewhat simple to perceive. Tracking down objects in the scene can ordinarily assist you with bettering comprehending human movement since it can give convenient data about the recent development. The vast majority of crafted by perceiving human movement includes human-situated scenes in an unfilled foundation where the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

entertainer is free to play out the movement. The improvement of a completely computerized human movement acknowledgment framework that can order the exercises of individuals with few blunders is an incomplete blockage, foundation clog, scale, viewpoint, lighting and appearance changes, and edges. What's more, the job of comment conduct is tedious and requires information on a specific occasion. What're more, likenesses between classes entangle the issue. That is, exercises in a comparative class can be tended to by different people with different body improvements, and exercises between different classes can be addressed by similar information that is difficult to perceive. The way that how an individual plays out an action relies upon his propensities, which leads to impressive issues in recognizing the primary 7 action. Another test is to construct visual models for ongoing preparing and examination of human developments utilizing lacking fundamental assessment datasets.

II. REALATED WORK

In the latest advancement a typical modal takes image or video as input and gives information about their pose and activity as output. Existing work done by different approaches can be compared on different basis such as number of activities estimated, model used etc. in the existing systems. We have compared 6 different papers for comparison between various activity recognition algorithms. Based on this comparison we came to conclusion that pose based approach is good and further we are using this approach through open pose to train our modal.

Pose based method uses image as input generates keypoints and matches posture variables #. In the technique dependent on picture structure, the stance's portrayal is considered as element to the grouping of the activity. [2] identify the individual in the frame and detect the coordinate information of the individual's body keypoints. [6] tackled the issue of labelling the video as per human activity which requires high computational requirement.

Apart from this single person activity recognition can be done using other approaches as well such as wearable sensors, radio frequency based, etc.; wearable sensors that uses sensors attached to human body using devices such as accelerometer and gyroscope to identify activity. Also uses different machine learning algorithms such as Logistic Regression, KNN, Random Forest etc. and detect different types of activities.

[4] uses sensors for identifying human activity. Accelerometer and gyroscope sensors are used for recognizing human activity. The fusion of both sensors provides significant results.

[1] propose a methodology for human pose estimation through walls and occlusion. The work is intended to manage the major issue of human pose estimation i.e., occlusion.

[5] the author demonstrates a deep learning approach to solve activity recognition problem. The authors divided their work into two parts: Pose estimation and Classification of activities.

Based on the study we done till now a brief description of the various methodology of HAR system is presented. Table 1 shows the comparative study performed on some selected paper between 2015 – present.

Table 1: Comparative Study

Ref. No	Author and Year	Dataset	Model	Activities	Accuracy
[1]	Mingmin Zhao et al., 2019	Self-Made Dataset	Visible Scene	Walking	70.7 %
			Through Wall		66.1 %
[2]	Gatt et al., 2019	COCO Dataset	Pre-Trained Model of PoseNet and OpenPose	Fall Detection	93 %
[3]	Ghazal et al., 2018	Own Dataset	CNN and SVM	Sitting	95.2 %
[4]	W. Jiang and Z. Yin, 2015	USC, UCI, SHO	DCNN and SVM	Walking, Sitting, Laying, Standing	97.89 %

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

III. PROPOSED APPROACH

The objective of our proposed work is to pose estimation and then use that extracted key point to classify different activities. The Initial step deals with the estimation of key point in an input image and then applying different supervised machine learning classifiers to classify the activity through the points. For classification purposes, we use Logistic regression, support vector machine, decision tree, etc. Figure 1 shows the architecture of the proposed approach.

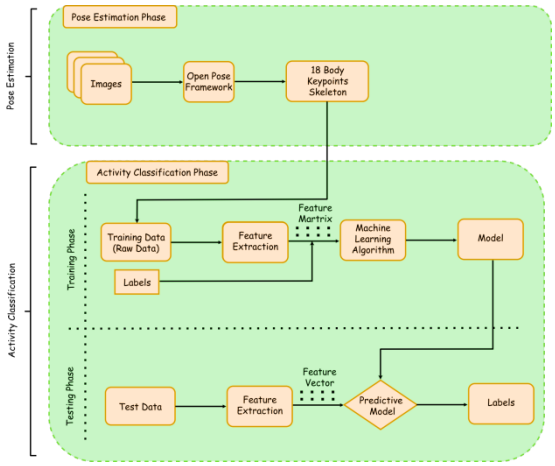

Figure 1: Proposed Architecture

1) *Pose Estimation*

Pose estimation is a technique of finding the location of the body joints of a person in an image or video. Once all the required body joints are extracted it uses them to construct a 2D skeleton by associating these joints. This work uses the Open pose Library to find all the required body joint/keypoints in an image.

The working of Open pose started with sending an image input to the convolution neural network (CNN) to extract the required feature. These features are then forwarded to the multi-stage CNN layers which generate the Confidence Map and Part Affinity Fields (PAF). Confidence maps store the information related to the key point location whereas PAF associates these joints, and Bipartite graphs help in capturing the individual in an image. Figure 2 shows the OpenPose pipeline.

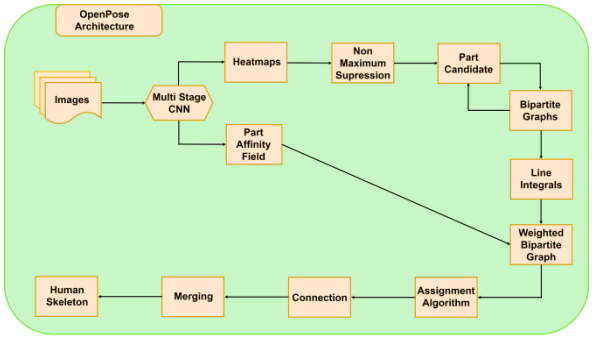

Figure 2: OpenPose Pipeline

1) *Confidence Map*

A Confidence Map is a 2D representation of the conviction that a specific body part can be situated in some random pixel.

Let J be the number of joints. Then, a confidence map is given by:

$$\text{the set } S = (S1, S2, S3, ..., Sj), \text{ where } Sj \in R^{w \times h},$$
$$j \in 1...J$$

2) *Part Affinity Fields (PAF):*

A Part Affinity field (PAF) is a bunch of stream handles that encode unstructured connections two by two between parts of a body. All aspects of the body have (PAF), for example neck, nose, elbows, and so forth. Let K be the number of pairs of body parts. Then, at that point, PAF are:

$$\text{the set } L = (L1, L2, L3, ..., Lk), \text{ where}$$
$$Lk \in R^{w \times h \times 2}, k \in 1...C$$

If a pixel is on a limb (body part), the value of Lk in that pixel is a 2D single vector from the initial joint to the final joint.

3) *Activity Classification:*

We figure out the action classification issue as a multiclass classification issue, which can be displayed utilizing different classification algorithms. The algorithm takes 18 body key points (both x and y coordinate of joints) as input for our model's training and testing. We used a supervised machine learning algorithm for classification as our dataset contains coordinates of joints with target activity. We used 6 different classifiers among the decision tree and random forest gives the highest precision.

4) *Dataset*

The most basic piece of the execution of a machine learning model is gathering a nice and adjusted dataset. There are numerous datasets accessible on the web yet for our need, we made our little dataset which comprises not many pictures. In order to perform the pose estimation, we need images for different activities. For our needs we consider only four activities i.e., standing, sitting, dancing and running. Our dataset comprises 1000+ images which we have downloaded online and some of them are clicked by smartphones. As the images are of different sizes so preprocessing is done to make them of the same size of 432 X 368 pixels. We categorize those images into 4 categories and apply pose estimation on them in order to find key points.

IV. ERROR EVALUATION

In the section we evaluate the approximate average error during pose estimation. To calculate this error between actual and estimated joints we have used the Euclidean Distance Formula which shows an average approximate error of 7.16673.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

Mathematically, it is defined by:

$$\sqrt{(Est.X - Act.X)^2 + (Est.Y - Act.Y)^2}$$

Figure 3 and 4 shows the deviation in X coordinate and deviation in Y coordinate, where the blue colored line represents estimated key points while the red line represents actual key points that must come.

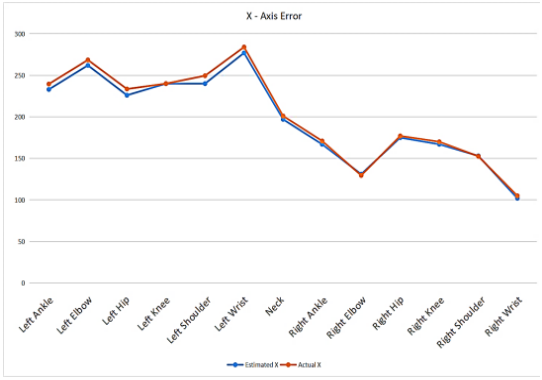

Figure 3: Deviation in x-coordinate

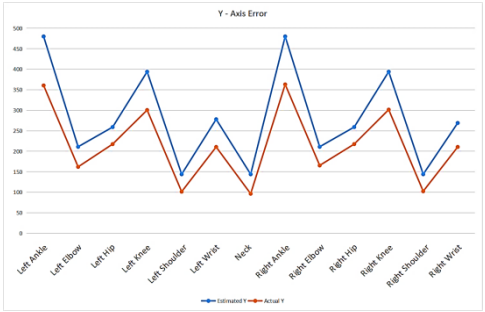

Figure 4: Deviation in y-coordinate

V. EXPERIMENTAL RESULTS

The accompanying five exercises are thought of for pose estimation and activity recognition and order: sitting, standing, dancing and running.

The experiments are conducted in Google Collab which is consist of Processor 1x Single core hyper threaded Xeon Processors @2.3Ghz i.e. (1 core, 2 threads) with RAM of 13 GB along with 100 GB disk space with GPU NVIDIA Tesla K80 compute 3.7, having 2496 CUDA cores, 12GB GDDR5 VRAM. These calculations are portrayed beneath with their confusion matrix. The presentation results are given in Table 1, which shows the review, accuracy of different classifiers utilized in the proposed approach.

1) Confusion Matrix

It is a two-dimensional lattice used to gauge the general exhibition of the AI learning calculation. In the grid, each row is related with the anticipated action class, and every segment is related with the genuine action class. The lattice contrasts the objective action and the movement anticipated by the model. This gives a superior thought of what kinds of blunders our classifier has made.

2) Classification Algorithm

a) Logistic Regression

It is one of the most eminent Machine Learning assessments, which submerge as the Supervised Learning system. It is like the Linear Regression (LR) aside from that how they are utilized. Linear Regression is utilized for managing Regression issues, while it is utilized for dealing with the solicitation issues. It predicts the yield of a full-scale subordinate variable. In this way, the result should be a straight out or discrete worth. It will overall be either Yes or No, 0 or 1, significant or False, and so on at any rate rather than giving the specific worth as 0 and 1, it gives the probabilistic qualities which lie some spot in the extent of 0 and 1.

The confusion matrix of this work is:

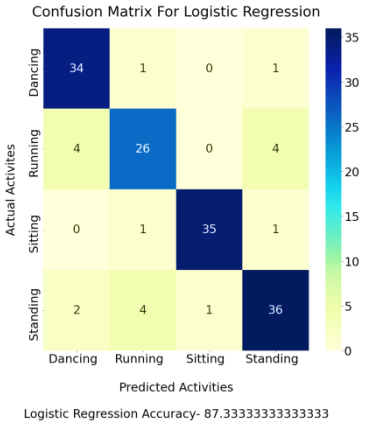

Figure 5: Confusion Matrix for Logistic Regression

b) K-Nearest Neighbors

It is one of the most un-complex Machine Learning estimations reliant upon administered learning methodology. Its estimation can be used for Regression similarly with respect to characterization anyway generally it is used for the order issues. Despite this effortlessness, we got extremely serious outcomes that are one justification behind utilizing this calculation in our work. We utilized various qualities for k and got the most noteworthy exactness in k = 5. In order to find the nearest neighbor, it uses Euclidean Distance between points. Figure 6 shows confusion matric for KNN.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

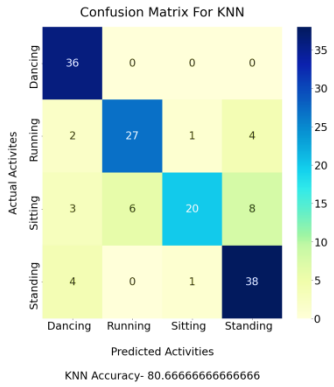

Figure 6: Confusion Matrix for KNN

c) Support Vector Machines

SVMs are perhaps the most notable Supervised learning calculation and are used for gathering and backslide issues. In any case, it is basically used for AI game plan issues. The justification behind the SVM estimation is to make ideal lines or plan restricts that can isolate the n-dimensional space into classes so new data centers can be easily situated in the right order later on. This breaking point for the best arrangement is known as the hyperplane. Confusion Matrix for same is displayed beneath:

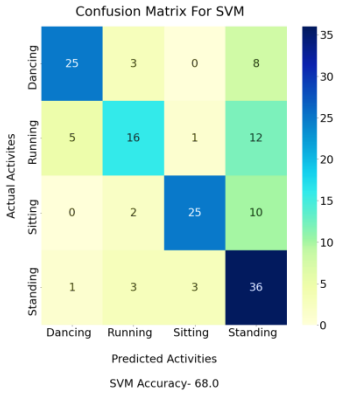

Figure 7: Confusion Matric for SVM

d) Random Forest

It is a classifier that contains distinctive choice trees on different subsets of the given dataset and takes the average to work on the discerning accuracy of that dataset." Instead of depending upon one choice tree, the irregular backcountry takes the figure from each tree and considering the greater part votes of suspicions, and it predicts the last yield. It is a managed learning computation used for request and backslide issues in AI. Confusion grid for same is displayed underneath:

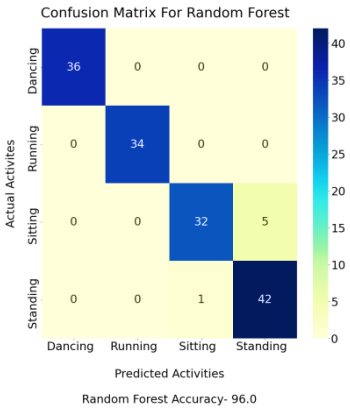

Figure 8: Confusion Matrix for Random Forest

e) Decision Tree

A Supervised learning technique can be utilized for both social event and Regression issues, however generally it is cherished for dealing with Classification issues. It is a tree-composed classifier, where inside focus focuses address the elements of a dataset, branches address the choice guidelines and each leaf place point keeps an eye on the result. A decision tree essentially addresses a solicitation, and thinking about the fitting response (Yes/No), it further split the tree into subtrees. Disarray network for this is given beneath:

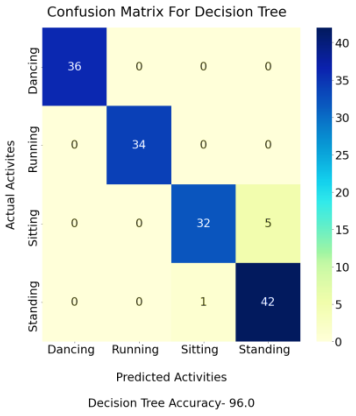

Figure 9: Confusion Matrix for Decision Tree

f) Naive Bayes

It is a supervised learning calculation that depends upon Bayes hypothesis and is used for managing plan issues. It is fundamentally used in a text strategy that goes with a high-dimensional planning dataset. Naive Bayes Classifier is one among the clear and best Classification assessments that associates in building the short AI models which will manufacture speedy suppositions. Confusion Matrix for this proposed work is displayed underneath:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

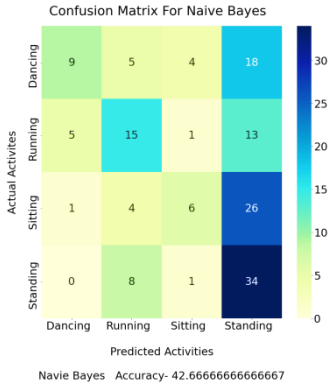

Figure 10: Confusion Matrix for Naïve Bayes

3) Accuracy Result

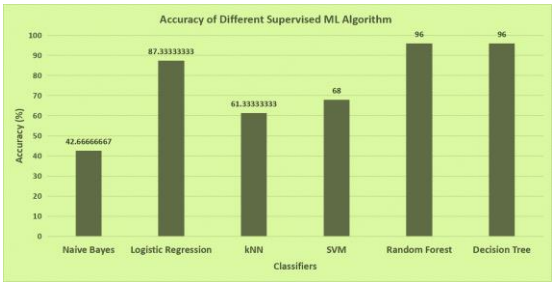

Figure 11: Accuracy of Different Algorithms

Figure 11 contains the accuracy comparisons between the different algorithms and Random Forest and Decision Tree performs very well and Naive Bayes fails in prediction.

CONCLUSION

This study presents a pose estimated based human activity recognition approach. This approach uses OpenPose Library to estimate the pose of a person in an image and for evaluating the error found during pose estimation it uses Euclidian distance to find the distance between the estimated joint and actual joint. After finding the coordinates of the joints it applies a different supervised machine learning classifier to classify the activities into four different categories viz. dancing, sitting, standing and running. For the requirement of data, we prepare our dataset which contains more than thousand images of four different activities. To deal with activity classification problems we used six classification algorithms (Logistic Regression, Decision tree, SVM, KNN, Random Forest, Naïve Bayes). After the experiment, we have found that Decision Tree and Random Forest produce the highest accuracy of 96.0% whereas Naïve Bayes produces the least accuracy of 42.66%. Although we are getting good results, this approach has some limitations such as, Occlusion, single person activity recognition.

REFERENCES

[1] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba and D. Katabi, "Through-Wall Human Pose Estimation Using Radio Signals," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.

[2] T. Gatt, D. Seychell and A. Dingli, "Detecting human abnormal behaviour through a video generated model," in 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 2019.

[3] U. S. K. Sumaira Ghazal, "Human Posture Classification using Skeleton Information," in *2018 International Conference on Computing Mathematics and Engineering Techonologies (iCoMET)*, Sukkur, Pakistan, 2018.

[4] W. Jiang and Z. Yin, "Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks," in MM '15: Proceedings of the 23rd ACM international conference on Multimedia, 2015.

[5] Amy L. Bearman, Stanford, Catherine Dong, "Human Pose Estimation and Activity Classification Using Convolutional Neural Networks (2015)".

[6] G. Shamsipour, J. Shanbehzadeh and A. Sarrafzadeh, "Human Action Recognition by Conceptual Features," in International MultiConference of Engineers and Computer Scientists 2017 (IMECS2017), Hong Kong, China, 2017.

[7] Tina, A. K. Sharma, S. Tomar and K. Gupta, "Various Approaches of Human Activity Recognition: A Review," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1668-1676, doi: 10.1109/ICCMC51019.2021.9418226.

[8] A. Gupta , K. Gupta, K. Gupta and K. Gupta , "A Survey on Human Activity Recognition and Classification," in 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020.

[9] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 43, Issue: 1, Jan. 1 2021), 2019.

[10] A. Singh, S. Agarwal, P. Nagrath, A. Saxena and N. Thakur, "Human Pose Estimation Using Convolutional Neural Networks," in 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019.

[11] A. Gupta, K. Gupta, K. Gupta, & K. Gupta "*Human Activity Recognition Using Pose Estimation and Machine Learning Algorithm*" in 2021 ISIC,2021.

[12] Fritz AI, "Pose Estimation Guide" [Online]. Available: https://www.fritz.ai/pose-estimation/

[13] B. Raj, "An Overview of Human Pose Estimation with Deep Learning," Apr 28, 2019 [Online]. Available: https://medium.com/beyondminds/an-overview-of-human-pose-estimation-with-deep-learning-d49eb656739b.

★★★