YOG-GURU: REAL-TIME YOGA POSE CORRECTION SYSTEM USING DEEP LEARNING METHODS

Ajay Chaudhari *, Omkar Dalvi †, Onkar Ramade‡, Prof. Dayanand Ambawade§
* † ‡ § § Department of Electronics and Telecommunications, Sardar Patel Institute of Technology, Mumbai, India
*ajay.chaudhari@spit.ac.in , †omkar.dalvi@spit.ac.in, ‡onkar.ramade@spit.ac.in ,§dd_ambawade@spit.ac.in

Abstract— **Self-learning is an integral part in Yoga, but incorrect posture while performing yoga can lead to serious harm to muscles and ligaments of the body. Thus to prevent this we present an intuitive approach based on machine learning techniques to correct the practitioner's pose while performing various yoga asanas. The proposed system is aimed at providing concise feedback to the practitioner so they are able perform yoga poses correctly and assist them in identifying the incorrect poses and suggest a proper feedback for improvement in order to prevent injuries as well as increase their knowledge of a particular yoga pose. A data-set of Five Yoga pose (i.e. Natarajasana and Trikonasana, and Vrikshasana and Virbhadrasana 1 & 2 and Utkatasana) has been created from collecting images from the Internet as well as from different individuals that took part in development of this system.**

A deep learning model is proposed which uses convolutional neural networks (CNN) for yoga pose identification along with a human joints localization model followed by a process for identification of errors in the pose for developing the system. Using the proposed system we have been able to achieve a classification accuracy of 95% for pose identification. After obtaining all the information about the pose of the user the system gives feedback to improve or correct the posture of the user.

Keywords - Human pose estimation, activity recognition, computer vision, real-time pose detection, yoga.

I. INTRODUCTION

Activity recognition is an emerging, significant as well as demanding topic among scholars and researchers around the globe in the field of computer vision and image processing [1]. A key reason for such an increase in demand is that the outcome of this activity recognition is becoming popular as well as useful in practical applications in the domain of human computer interaction, healthcare and to some extent in sports.

The ancient Indian art of connecting mind and body, Yoga, is a great way to exercise which helps in maintaining mental clarity and calmness. Yoga is a way to unify the physical functioning of the body and the brain, which promotes our physical well-being and also boosts the mental state, thus it is gaining popularity nowadays as it was popular in the ancient times [2]. However, when practicing yoga, the person must follow the proper guidelines to gain maximum benefit from following a particular yoga routine. Hence it is necessary for practitioner to invest their time exercising by themselves along with the regular training courses given by an instructor. The importance of such system comes into picture when the practitioner is trying to learn yoga by themselves as they are not instructed by a professional, so they tend to make slow progress initially and can even hurt themselves during self-training because of incorrect practices.The practitioner have to follow certain steps and guidelines for taking maximum benefit from a particular pose and in lot of cases failing to do so may lead to serious harm due to incorrect postures. If these incorrect practices are continued for a prolonged period it may lead to long-term pain in the joint [3]. This became our prime motivation for developing a yoga self-training systems for practitioner to exercise on their own with the help of a mobile or web based application only, so that proper instructions can be delivered to help them in very convenient manner so that they can adjust their poses and learning about their incorrectness in real time and there is also an increasing demand for the development of computer-assisted training systems from the practitioner side to help them to improve their knowledge and understanding about different types of yoga poses and also protect them from injury that might occur during learning phases.

These types of applications also comes handy in the period of lockdown, like the lockdown we have seen in the 2020 due to the coronavirus pandemic, where the movement of people is restricted to a lot of extent and they can use such applications very conveniently at their home also.

II. LITERATURE SURVEY

There have been multiple methods that have been used to perform human pose estimation. For our case we are going to focus on single person pose estimation.The traditional approach involves performing calculation over different combinations of local parameters of body parts and the relative dependencies between them. Two major methods are used, the tree structured model and non- tree based models. The tree based model uses parametric encoding to find the spatial relation between related body parts using kinematic chain links. While the non-tree models use edge capture method, occlusion, symmetry and link relationship to obtain reliable observation related to body parts.

One of the ways to capture the posture of individual include use of depth cameras based and kinetic sensors.Yu-Liang Hsu et al in 2018 developed a inertial sensor network that worked on activity recognition algorithm that could accurately predict the daily human activity and sports activities. [1].They proposed a method of using 2 inertial sensing devices which comprised of a micro-controller, a tri-axial accelerometer and gyroscope.They have used the non-parametric weight feature extraction (NWFE) with principal component analysis(PCA) for reducing the feature dimension of inertial signal. In 2018 Chen HT et al proposed a system of tracking human pose using Kinetic console sensors in "Computer-assisted yoga training system" [2].In this method they have suggested the used of feature based approach for designing a yoga training system. Kinect's inertial sensors are used for extracting the user's body contour. They are then used to capture the body map and extract posture information like the dominant axis, skeleton and contour based feature points. In 2017 Ya Lui et al proposed using a interacting multiple model (IMM) based on unscented Kalman filters [4].The IMM algorithm is capable of adaptive infusing the posture of models and generating optimally estimation via flexible model weighting. Another method was proposed by Siddharth Sreeni et al in "Multi-Modal Posture Recognition System for Healthcare Applications" [5] here they have used a combination of 3D depth mapping technique in combination with sensor based inertial measurements with nine degrees of freedom. These technique are used to capture the posture parameters of the person in real time and then to predict their pose it is compared with existing data parameters of known poses.

The drawback of these methods are that depth sensors are generally not available to users and can also cause interference while performing poses.

Other methods include contour based pose estimation techniques. In "Yoga tutor: visualization and analysis using SURF algorithm" [6] a SURF algorithm approach is taken to identify the pose of users in images.SURF (Speeded Up Robust Feature) algorithm is an advanced version of the SIFT algorithm.The approach includes gray scaling i.e converting the image into gray scale, thresholding to eliminate the background and separate it from the user.Erosion and dilation to increase the smoothing the boundary noise of the image. Further the Canny algorithm is applied in the prefinal stage which does smoothing, gradient and edge tracking. And finally the SURF algorithm is applied to compare the reference video with the practitioner video. In 2017 Soltow.E et al proposed a model based pose estimation technique that used segmentation of contours in X ray images. Pose estimation is done using Fourier descriptors [7].

Another method of pose estimation is through neutral network and deep learning. In "Human Pose Estimation Using Convolutional Neural Networks" the authors have created a simple model using convolutional neural network that estimates the poses and demonstrates the potential of CNN's [8].The model predicts a set of locations and a set of confidence maps with body part location and degree of association between parts in the form of 2D vector fields.

In [9] "Automated daily human activity recognition for video surveillance using neural network" the authors have used a series of digital image processing like background subtraction, binarization,and morphological operation.Based on the activities features database a robust neural network was built which was extracted from the frame sequence. A Multi-layer feed forward perceptron network is used to classify the activities model based on the data set.

In 2014 Alexander Toshev and Christian Szegedy presented the paper "DeepPose: Human Pose Estimation via Deep Neural Networks" [10] in which they proposed a method for pose detection using Deep Neural Network(DNN). They considered pose identification as a joint regression problem and depicted how to successfully cast it in DNN model. The location of each body point is regressed using the full image as input for a generic convolutional DNN with 7 layers.There are two major advantages to using this method. The first is that we are able to fully capture the context of each body point and their relation with adjacent joints. The Second is that this approach is much faster and easier to formulated methods based on graphical models.The only issue with this method that it faces from multiple localization problems with relation to the accurate position of points in non ideal images.Another example of image based pose identification is shown by Mohanty A et al in ”Robust pose recognition using deep learning” [11] here they have discussed two deep learning methodologies to tackle the problem.The first is convolutional neural network(CNN) and the second is stacked auto encoder(SAE). Using these two techniques they have been able to achieve high prediction and recognition percentage.

III. Proposed Methodology

The approach seeks to extract the user's body posture automatically and compare it with the predefined collection of acceptable yoga postures using 15 key body point monitoring on the user's real time camera feed performing the asanas. The procedure consists of 4 stages, which are key point extraction, pose identification and error estimation and feedback. In the initial stage we run the key point extraction model on the real time video of the user performing the yoga asanas to extract the body posture of the user. In the next stage the model predicts the asana being done by the user and in the final stage, visual feedback is provided to the user instructing him to make changes in his posture after being compared with the ideal one.

A. Data Collection

Convolutional Neural Network (CNN) performs impressively in computer vision problems such as image classification, object detection, etc especially when the size of data-set used for training the model is of large size.

Our first step towards identifying the incorrection in a particular pose is to identify the pose itself which is done via a CNN classifier for different yoga poses.

Hence, to solve this problem, we have used the concept of pose classification using deep learning techniques, instead of producing key points and skeleton annotations for pose

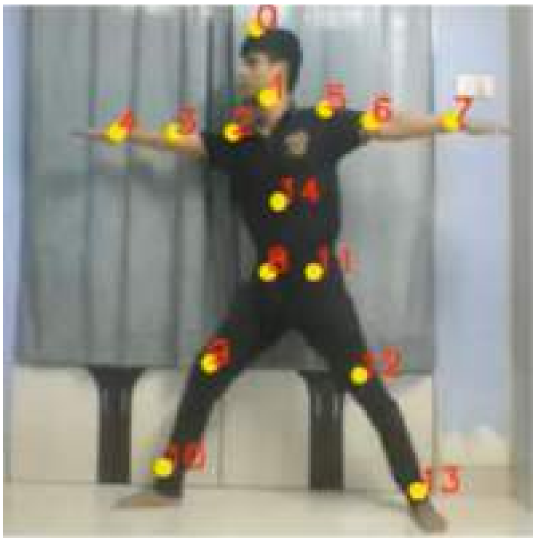

Fig. 1. 15-key-points detected by the model.

identification which may not be possible due to heavy computational requirement, we used a CNN classifier for this step. The model was trained with data acquired from the Yoga-82 dataset [12], which consists of 28478 images of 82 different yoga poses.

B. Pose Extraction

In the first step, the device camera captures the user's real-time video, which is then processed by the model to extract the 15 main body points from the stream, as shown in figure 1, from the human body 's joints. For that particular point, each of the 15 joint points is labelled with the index value along with the x and y co-ordinates that are unique to each joint. These coordinates provide the value of the joint in the x and y directions.

C. Pose Identification

After extracting the key points of the human skeleton along with their co-ordinates, the next step is identifying the pose of the user. We use all the 15 points for pose recognition(Fig. 1). In that particular case, the x and y coordinates of each point are used to determine the structure of the human body, which is then compared against the structure of each asana 's ideal poses previously fed to the model. The model outputs the predicted asana with the accuracy rating over the user image after recognition as shown in Fig. 2.

D. Error Estimation and Feedback

Finally , the model compares the derived key points of the user image to the predefined set of reference key points of that asana 's ideal body structure after successfully identifying the pose of the user. The location of each key point with respect to that of its neighbouring key point is tested and, if any mistake or inconsistency is detected, a text message appears that directs the user to make any required adjustments to the current pose to correct the error. In Fig. 3, the model has classified the image output in the first row as correct, while those in the second row have been correctly classified as erroneous, and the user must correct his error in the performed pose.

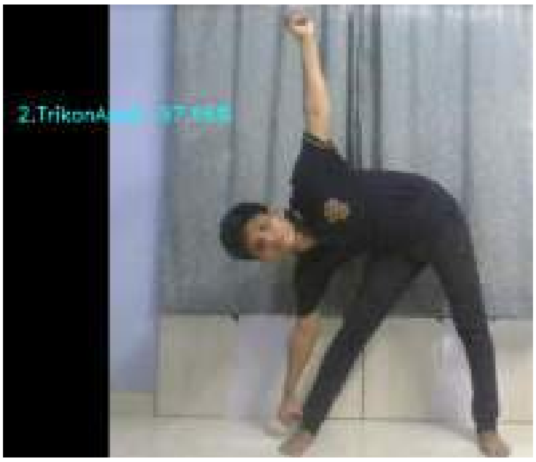

Fig. 2. Predicted Asana by the model.

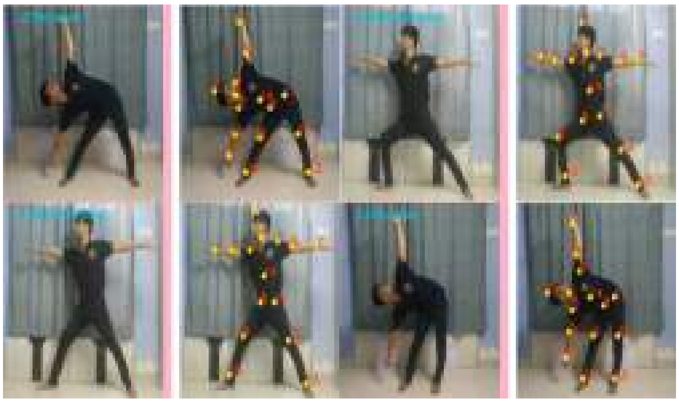

Fig. 3. Model output for various poses

IV. SYSTEM ARCHITECTURE

The proposed system is capable of recognizing six yoga asanas in real time and from pre-recorded videos too. These asanas include Natarajasana, Vrikshasana, Trikonasana, Virbhadrasana 1 & 2, Utkatasana. System architecture diagram is shown in the Fig. 4.

As discussed in the methodology, the flowchart in Fig. 5 shows the working of the above proposed system. The First part is data collection, which can be done by either a process for collecting the frames from the videos which is running in parallel with detection or can be from pre-recorded videos. Second part is the pose extraction,a key-point detection model is used to identify the joint locations using Part Confidence Maps and Part Affinity Fields. The detected key points are passed to our model where CNN identifies the pose and a predetermined mathematical model is used to identify in corrections in the pose.

- Step 1: We begin the process by capturing the video input of the user performing the yoga poses from the

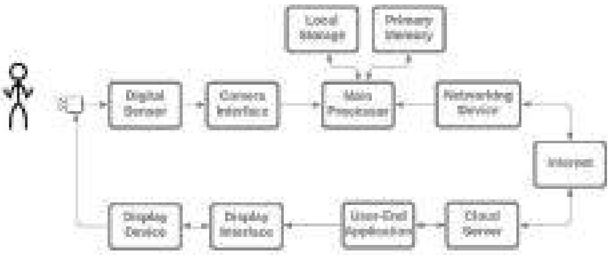

Fig. 4. Block diagram of the System

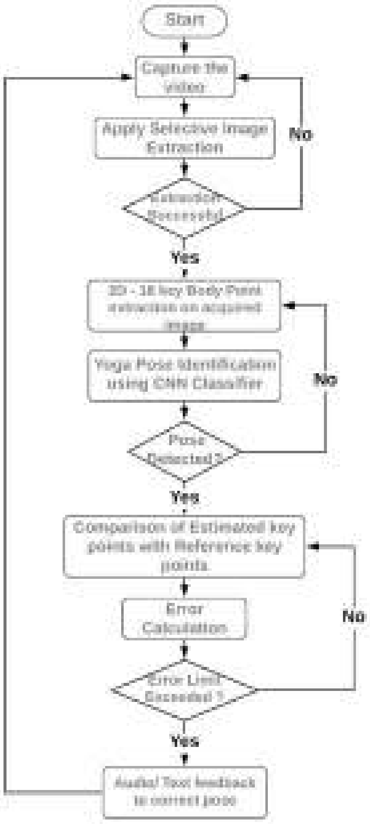
Start
Capture the video
Apply Selective Image Extraction
No
Extraction Successful
Yes
2D - 18 key Body Point extraction on acquired image
Yoga Pose Identification using CNN Classifier
No
Pose Detected ?
Yes
Comparison of Estimated key points with Reference key points
Error Calculation
No
Error Limit Exceeded ?
Yes
Audio/ Text feedback to correct pose

Fig. 5. Flow Chart of the System

camera. Then we apply Selective Image Extraction to get the clear image of the user and the background disturbances are removed.

- Step 2: Then a 2 dimensional 15 key body point extraction is applied on the image to procure the key points of the user's body. Then the image is passed through the Yoga Pose Identification CNN classifier, which detects the yoga pose the user is trying to perform.
- Step 3: If the pose is detected correctly then we compare the estimated key points with the reference key points of the predefined pose.
- Step 4: Then the error is calculated between the estimated and reference poses and necessary feedback is given to the user if the error exceeds the threshold limit

A. *Postural feature extraction Techniques*

Postural feature extraction is necessary in order to build a Yoga self-training system, one such method described in literature survey is by using a Kinect depth camera, however, it uses manual feature extraction and requires different model for different asana.

Human skeleton, a delegate feature, which are compulsory for describing different types of human posture. There are various techniques to obtain a skeleton of the human body that gives the location of different point on body whuch are then utilize to estimate the posture. Various techniques are mentioned in the literature, such as thinning and distance transformation. However, these techniques are computational very heavy and are not suitable for general mobile applications and are sensitive to noise as well. Hence these approaches are now being replaced with deep-learning based techniques since the arrival of DeepPose [10]. It estimates the movement of a person and determines the location of hidden body parts that are difficult to determine in normal scenarios.

An example of a postural feature extraction model would be OpenPose [13]. It is a real time human body key point detection model that is based on confidence mapping and greedy algorithm. OpenPose is able to detect key points for the human body , facial structure and feet from single frame images. Human pose detection is a fast developing field and OpenPose caused a major revolution by drastically reducing the computation time without sacrificing the accuracy of the detection model.They were able to provide enough efficiency for the model to be able to run on real time video input.Table I shows the list of all the body keypoints that the OpenPose model is able to detect.

Point No	Key Point Name	Point No	Key Point Name
0	Head	8	Right hip
1	Neck	9	Right knee
2	Right shoulder	10	Right foot
3	Right elbow	11	Left hip
4	Right wrist	12	Left knee
5	Left shoulder	13	Left foot
6	Left elbow	14	Torso
7	Left wrist		

TABLE I
15 KEY-POINT WITH NAMES

B. *Comparison Methodology for Error Identification*

As discussed earlier, that we have utilized a convolutional neural network (CNN) to obtain the 15 point on the body with the help of which a skeletons like figure of the user's pose is created as shown in the Fig. 6. At the same time we have created a target pose which acts as a reference pose for comparing the similarity of the user's pose. The reference pose is the desired yoga pose that the user is trying to perform. The target pose and pose acquired from the key-point detection model will be compared to check the similarity between various angle and joints.

This similarity will be used to measure the correctness of the user's pose. The technique to find the incorrectness is to calculate the angles between the joints of the user and based on the domain knowledge of yoga we check if the angles should be within the tolerance level for performing a yoga pose.

The point detection model returns the coordinates and the confidence level of the different body parts as output. The 15 point output format is used for this project which returns the x and y coordinates of 15 body parts along with their confidence levels. Our hypothesis is that the coordinates of the various body parts in a human body extracted from an image contains enough information to determine if a pose is performed correctly.

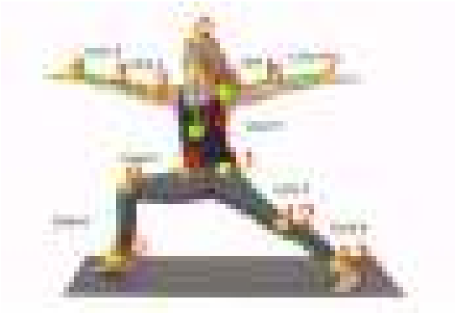

Fig. 6. Virbhradasana 2 Pose with 15 keypoints detected

Angle Of line	Between points	Angle w.r.t Horizontal
Right Shoulder (Line1)	2 & 3	20
Right elbow (Line2)	3 & 4	20
Left shoulder (Line3)	5 & 6	20
Left elbow (Line4)	6 & 7	20
Right knee (Line5)	8 & 9	20
Right Foot (Line6)	9 & 10	90
Left Knee (Line7)	11 & 12	45
Left Foot (Line8)	12 & 13	45
Torso (Line9)	1 & 14	10

TABLE II
IDEAL ANGLES FOR VIRBHRADASANA

The Table II shows the ideal angles between various joint of the body while performing Virbhradasana 2, with the addition of some amount of tolerance the error is calculated for the given pose in real time. Similarly other reference model for each is created and can be utilized in the similar fashion for error identification.

V. RESULTS

Table III shows the performance parameters for our model. These have been calculated for 6 different posses. The values calculated are precision, recall, F1-score. We have been able to achieve a minimum F1 score of of 0.72.

In Fig. 7 we can see the loss/accuracy per epoch cycle. Using technique of image augmentation we were able to add more information and improve the accuracy of our model and reached upto 95% of training accuracy as evident in the figure.

TABLE III
PERFORMANCE TABLE

No	Asana Names	Precision	Recall	F1-Score	Support
1.	Natrajasana	0.89	0.80	0.84	10
2.	Trikonasana	0.94	0.75	0.79	26
3.	Utkatasana	0.88	0.82	0.85	34
4.	Virbhadrasana 1	0.70	0.93	0.72	43
5.	Virbhadrasana 2	0.79	0.84	0.82	32
6.	Vrikshasana	0.89	0.69	0.78	36
	accuracy			0.81	181
	micro average	0.87	0.80	0.82	181
	weighted average	0.85	0.81	0.81	181

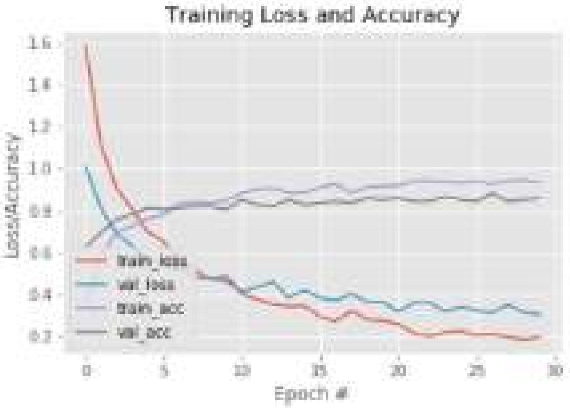

Fig. 7. Model loss-accuracy graph

In Fig. 8 we can see the dashboard of the YogGuru website. The dashboard is primarily divided into 3 portions. The first is the normal live feed of the user which is displayed on the right hand side.The second window is where the body points of the user are displayed. And the bottom window is there the suggestion for user to correct or improve their pose are currently displayed. In the future the suggestions will be given in the form of audio instead of text output.

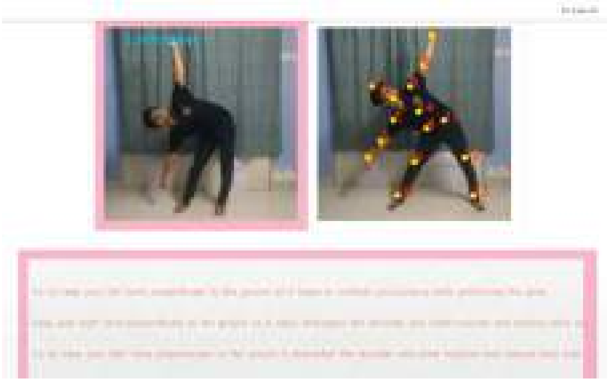

Fig. 8. Website Dashboard

VI. CONCLUSIONS

In this project, we have developed a system which consists of a pipeline for pose identification and then point localization on human body and then followed by a error identification process. This system aims at helping people to perform yoga in a correct manner on their own and prevent injuries that can happen due to incorrect practices in yoga.

Using deep learning techniques the system is able to analyse the pose of the user from the front view and give a feedback to them for improving their yoga pose. A user friendly dashboard was also created in this project on which the developed model is deployed.

The system delivers satisfactory results as shown in the above experimental analysis, but there are some areas for improvement as well, for example the pose identification model provides less accuracy for Virabhadrasana-1 due its similarity with other poses such as Virabhadrasana-2 and Utkatasana. This inaccuracy can be overcome by improving the methods for feature point detection or by redesigning the methods for features extraction which is essential for pose identification. Besides, the working of this system be enhanced by adding more modules of other yoga poses. The future scope for this project is to enhance it by adding voice feedback and well an including use-cases in other field's like sports, dance etc.

REFERENCES

[1] Y. Hsu, S. Yang, H. Chang and H. Lai,"Human Daily and Sport Activity Recognition Using a Wearable Inertial Sensor Network," *IEEE Access,*vol. 6, pp. 31715-31728, 2018.

[2] Chen, H., He, Y. Hsu, C. "Computer-assisted yoga training system" *M*ultimed Tools Appl 77, 23969–23991 (2018). https://doi.org/10.1007/s11042-018-5721-2

[3] N. Belling, "The Yoga Handbook: A Complete Step-by-Step Guide", *New Holland Puvlishers Ltd, 2001*

[4] Y. Liu, X. Tian and X. Xu,"Posture estimation system by IMM-based unscented Kalman filters," *2017 Chinese Automation Congress (CAC),*Jinan, 2017, pp. 2363-2368.

[5] Sreeni.S, R.H.S, R.H, V.S. (2018)," Multi-Modal Posture Recognition System for Healthcare Applications". *TENCON 2018 - 2018 IEEE Region 10 Conference*

[6] Patil, Siddharth Pawar, Amey Peshave, Aditya Ansari, Aamir Navada, Arundhati. (2011)." Yoga tutor visualization and analysis using SURF algorithm." *2011 IEEE Control and System Graduate Research Colloquium*

[7] Soltow E., Rosenhahn B.(2016)," Automatic Pose Estimation Using Contour Information from X-Ray Images," *Huang F., Sugimoto A. (eds) Image and Video Technology – PSIVT 2015 Workshops. PSIVT 2015. Lecture Notes in Computer Science,*vol 9555. Springer, Cham.

[8] Singh, A., Agarwal, S., Nagrath, P., Saxena, A., Thakur, N. (2019). "Human Pose Estimation Using Convolutional Neural Networks". *2019 Amity International Conference on Artificial Intelligence (AICAI)*

[9] Babiker, Mohanad Khalifa, Othman Htike, Kyaw Hashim, Aisha Zaharadeen, Muhamed. (2017). "Automated daily human activity recognition for video surveillance using neural network" *2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*

[10] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," *2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1653-1660, doi: 10.1109/CVPR.2014.214.*

[11] Mohanty A., Ahmed A., Goswami T., Das A., Vaishnavi P., Sahay R.R. (2017) "Robust Pose Recognition Using Deep Learning". *In: Raman B., Kumar S., Roy P., Sen D. (eds) Proceedings of International Conference on Computer Vision and Image Processing. Advances in Intelligent Systems and Computing, vol 460. Springer, Singapore. https://doi.org/10.1007/978-981-10-2107/7/9*

[12] M. Verma, S. Kumawat, Y. Nakashima and S. Raman, "Yoga-82: A New Dataset for Fine-grained Classification of Human Poses," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 2020, pp. 4472-4479, doi: 10.1109/CVPRW50498.2020.00527.*

[13] S. Qiao, Y. Wang and J. Li, "Real-time human gesture grading based on OpenPose," *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, 2017, pp. 1-6, doi: 10.1109/CISP-BMEI.2017.8301910.*