

Estimation of Yoga Postures Using Machine Learning Techniques

Abstract

Yoga is a traditional Indian way of keeping the mind and body fit, through physical postures (asanas), voluntarily regulated breathing (*pranayama*), meditation, and relaxation techniques. The recent pandemic has seen a huge surge in numbers of yoga practitioners, many practicing without proper guidance. This study was proposed to ease the work of such practitioners by implementing deep learning-based methods, which can estimate the correct pose performed by a practitioner. The study implemented this approach using four different deep learning architectures: EpipolarPose, OpenPose, PoseNet, and MediaPipe. These architectures were separately trained using the images obtained from S-VYASA Deemed to be University. This database had images for five commonly practiced yoga postures: tree pose, triangle pose, half-moon pose, mountain pose, and warrior pose. The use of this authentic database for training paved the way for the deployment of this model in real-time applications. The study also compared the estimation accuracy of all architectures and concluded that the MediaPipe architecture provides the best estimation accuracy.

Keywords: Artificial intelligence, deep learning, machine learning techniques, pose estimation techniques, skeleton and yoga

Introduction

Yoga is an ancient Indian science and a way of living that includes the adoption of specific bodily postures, breath regulation, meditation, and relaxation techniques practiced for health promotion and mental relaxation. In recent years, yoga has been adopted internationally for its health benefits. Among several techniques, physical postures have become very popular in the Western world. Yoga is not only about the orientation of the body parts but also emphasizes breathing and being mindful.^[1] The traditional Sanskrit name for Yoga postures is asanas. During the pandemic, many people have used yoga to keep themselves physically and mentally fit.^[2] Many people practice fine forms of asanas, without a teacher to guide them: either because no trained yoga instructors are available or due to unwillingness to engage one. Nevertheless, it is important to perform asanas correctly, so the practitioner does not sustain injury.^[3] Furthermore, asanas should be practiced systematically, paying attention to the orientation of the limbs and the breathing. Improper stretching or performing inappropriate

asanas and breathing inappropriately when exercising can be injurious to health. Improper postures can lead to severe pain and chronic problems.^[4] Hence, a scientific analysis of asana practice is all important. The present work was developed, with this in mind.

Pose estimation techniques can be used to identify the accurate performance of yoga postures.^[5] Pose estimation algorithms have been used to mark the key points and draw a skeleton on the human body for real-time images and used to determine the best algorithm for comparing the poses. Posture estimation tasks are challenging as they require creating datasets from which real-time postures can be estimated.^[6]

This study estimated the five asanas performed by the participant using four different deep learning architectures: EpipolarPose, OpenPose, PoseNet, and MediaPipe. These architectures are especially suitable for pose estimation. Deep learning architectures were trained for the abovementioned five asanas. The training was carried out on an authentic database at S-VYASA Deemed to be University, hence suitable for real-time and practical applications. The dataset consisted

**D. Mohan Kishore,
S. Bindu¹,
Nandi
Krishnamurthy
Manjunath**

Division of Yoga and Life Sciences, Swami Vivekananda Yoga Anusandhana Samsthana (S-VYASA), ¹Department of Electronics and Communication Engineering, B N M Institute of Technology, Bengaluru, Karnataka, India

Address for correspondence:

Mr. D. Mohan Kishore,
Swami Vivekananda Yoga
Anusandhana Samsthana
(S-VYASA), Jigani,
Bengaluru – 560105,
Karnataka, India.
E-mail: mohankishorejain@
svyasa.edu.in

Access this article online

Website: www.ijoy.org.in

DOI: 10.4103/ijoy.ijoy_97_22

Quick Response Code:



How to cite this article: Kishore DM, Bindu S, Manjunath NK. Estimation of yoga poses using machine learning techniques. *Int J Yoga* 2022;15:137-43.

Submitted: 29-May-2022

Revised: 13-Jun-2022

Accepted: 15-Jun-2022

Published: 05-Sep-2022

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

of about 6000 images of the above five postures, of which 75% of the dataset was used in training the model, whereas 25% was used for testing.

Human Body Modeling

Human body modeling is essential to estimate a human pose by locating the joints in the body skeleton from an image. Most methods use kinematic models where the body's kinematic structure and shape information is represented by its joints and limbs.^[7] Different types of human body modeling are shown in Figure 1.

The human body can be modeled using a skeleton-based (kinematic) model, a planar (contour-based) model, or a volumetric model, as shown in Figure 1. The skeleton-based model represents a human body having different key points showing the positions of the limb with orientations of the body parts.^[8,9]

However, the skeleton-based model does not represent the texture or shape of the body. The planar model represents the human body by multiple rectangular boxes yielding a body outline showing the shape of a human body.^[10] The volumetric model represents a three-dimensional (3D) model of well-articulated human body shapes and poses.^[11] The challenges involved in human pose estimation are that the joint positions could change due to diverse forms of clothes, viewing angles, background contexts, and variations in lighting and weather,^[12] making it a challenge for image processing models to identify the joint coordinates and especially difficult to track small and scarcely visible joints.

Human Pose Estimation

Computer vision is used to estimate the human pose by identifying human joints as key points in images or videos, for example, the left shoulder, right knee, elbows, and wrist.^[13] Pose estimation tries to seek an exact pose in the space of all performed poses. It can be done by single pose or multipose estimation: a single object is estimated by the single pose estimation method, and multiple objects are estimated by multipose estimation.^[14] Human posture

assessment can be done by mathematical estimation called generative strategies, also pictorially named discriminative strategies.^[15] Image processing techniques use AI-based models, such as convolutional neural networks (CNNs) which can tailor the architecture suitable for human pose inference.^[16] An approach for pose estimation can be done either by bottom-up/top-down methods.

In the bottom-up approach, body joints are first estimated and then grouped to form unique poses, whereas top-down methods first detect a boundary box and only then estimate body joints.^[17]

Pose estimation with deep learning

Deep learning solutions have shown better performance than classical computer vision methods in object detection. Therefore, deep learning techniques offer significant improvements in pose estimation.^[18,19]

The pose estimation methods compared in this research include EpipolarPose, OpenPose, PoseNet, and MediaPipe.

EpipolarPose

The EpipolarPose constructs a 3D structure from a 2D image of a human pose. The main advantage of this architecture is that it does not require any ground truth data.^[20] A 2D image of the human pose is first captured, and then an epipolar geometry is utilized to train a 3D pose estimator.^[21] Its main disadvantage is requiring at least two cameras. The sequence of the steps for training is shown in Figure 2. The upper row of the Figure 2 (orange) depicts the inference pipeline and the bottom row (blue) shows the training pipeline.

The input block consists of the images of the same scene (human pose) captured from two or more cameras. These images are then fed to a CNN pose estimator. The same set of images are then fed to the training pipeline, and after triangulation, the 3D human pose obtained (V) is fed back to the upper branch. Hence, this architecture is self-supervised.

OpenPose

The OpenPose is another 2D approach for pose estimation.^[22] The OpenPose architecture is shown in Figure 3a-c. Input images can also be sourced from a webcam or CCTV footage. The advantage of OpenPose is the simultaneous detection of body, facial, and limb key points.^[23] Figure 3a shows VGG-19, a trained CNN architecture from the Visual Geometry Group. It is used to classify images using deep learning. It has 16 convolutional layers along with 3 fully connected layers, altogether making 19 layers and the so-called VGG-19. The image extract of VGG-19 is fed to a "two-branch multistage CNN," as shown in Figure 3b. The top part of Figure 3c predicts the position of the body parts, and the bottom part represents the prediction of affinity fields,

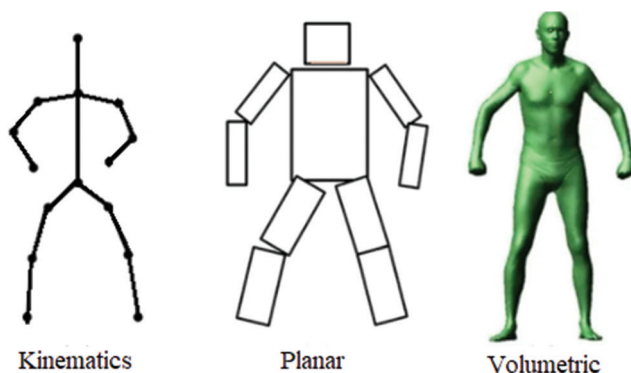


Figure 1: Human body modeling

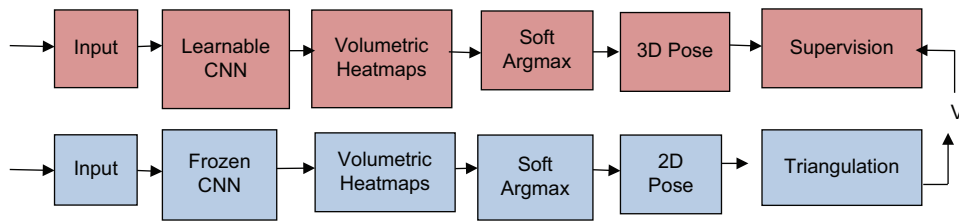


Figure 2: The architecture of the EpipolarPose involved during training

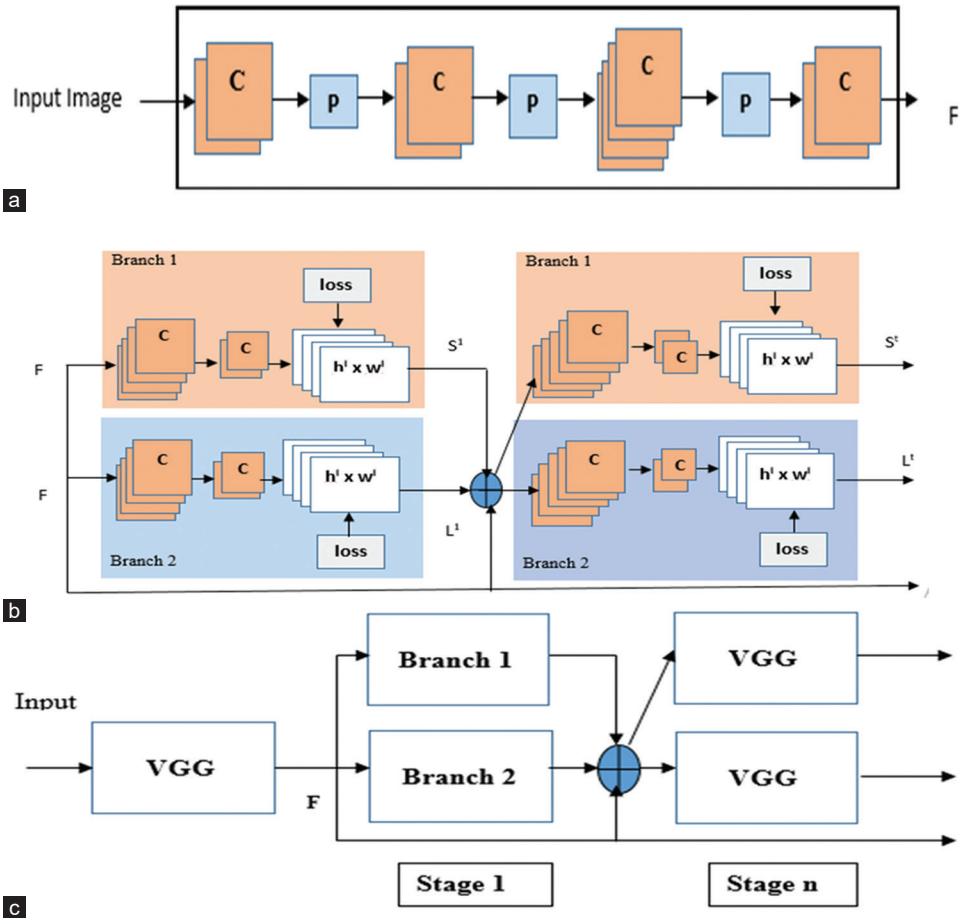


Figure 3: (a) VGG-19 Convolution Neural Network (C-Convolution, P-Pooling). (b) Convolution layer branches. (c) OpenPose architecture

i.e., the degree of association between different body parts. By these means, the human skeletons are evaluated in the image.

PoseNet

The PoseNet can also take video inputs for pose estimation; it is invariant to image size; hence, it gives a correct estimation even if the image is expanded or contracted^[24,25] and can also estimate single or multiple poses.^[26] The architecture shown in Figure 4 has several layers with each layer having multiple units. The first layer includes input images to be analyzed; the architecture consists of encoders that generate visual vectors from the image. These are then mapped onto a localization feature vector. Finally, two separated regression layers give the estimated pose.

MediaPipe

This is an architecture for reliable pose estimation. It takes a color image and pinpoints 33 key points on the image. The architecture is shown in Figure 5.

A two-step detector-tracker ML pipeline is used for pose estimation.^[27] Using a detector, this pipeline first locates the pose region-of-interest (ROI) within the frame. The tracker subsequently predicts all 33 pose key points from this ROI.^[28]

Methodology Adopted

Initially, the image of a yoga practitioner performing an asana was captured by a camera and fed separately to the four deep learning architectures, which then estimate the

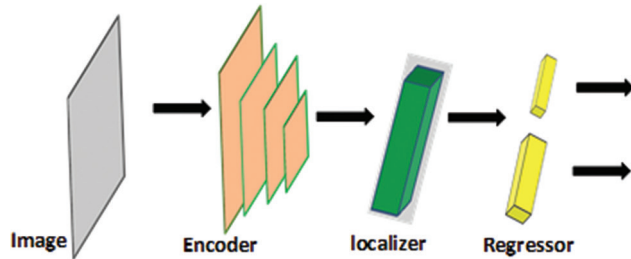


Figure 4: PoseNet architecture

pose performed by the practitioner by comparing it with the pretrained model. If it does not match any of the five asanas, an error was shown.

Twenty practitioners in the age group of 18–60 years performing different postures in real time were captured and fed separately to the proposed architectures, and a comparison of the estimated accuracy was done.

Results

Pose estimation for five yoga postures was done using different proposed techniques. The results of pose estimation were shown for each of the five asanas for all the four architectures used. For simplicity, the images of the same individual were shown (after taking consent) for all estimations and comparisons. The five yoga poses considered for posture estimation are as follows:

- a. Ardha Chandrasana/half-moon pose,
- b. Tadasana/mountain pose,
- c. Trikonasana/triangular pose,
- d. Veerabhadrasana/warrior pose
- e. Vrukshasana/tree pose.

Results of pose estimation using EpipolarPose

The pose estimation results obtained for five yoga postures using an EpipolarPose are shown in Figure 6.

Results of pose estimation using OpenPose

The pose estimation results obtained for five yoga postures using OpenPose are shown in Figure 7.

Results of pose estimation using PoseNet

The pose estimation results obtained for five yoga postures using PoseNet are shown in Figure 8.

Results of pose estimation using MediaPipe

The pose estimation results obtained for five yoga postures using MediaPipe are shown in Figure 9.

Pose estimation of the five yoga postures was done for different methods, as shown in Figures 6-9. After validation of the model, 20 sample images were captured in real time and were fed individually to the model, and the posture accuracy was estimated. The average value of accuracy is summarized in Table 1. Here, the method used for

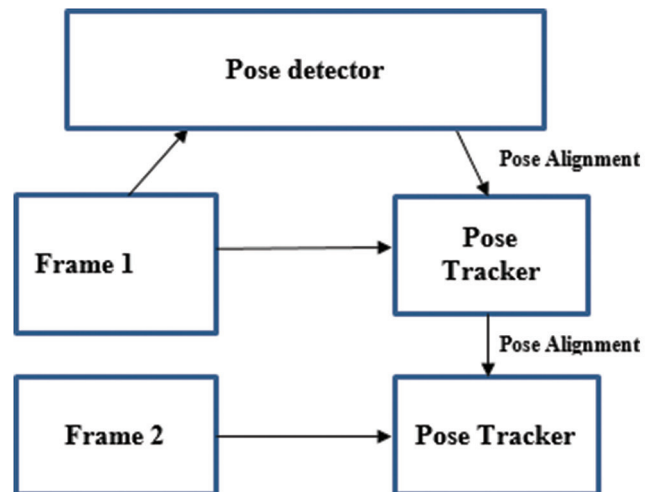


Figure 5: Human pose estimation pipeline overview

calculating the accuracy is the classification score, which is the ratio of the number of correct predictions (CP) made to the total number of predictions (TP) (i.e., total number of predictions = the sum of CP and the number of wrong predictions (WP))

$$TP = \frac{CP}{CP + WP}$$

It is observed that the accuracy of prediction using EpipolarPose was around 50%. This is because the EpipolarPose is generally suited for describing and analyzing multicamera vision systems dealing with two viewpoints of the same points in a pair of images.^[29] As this work involves capturing the image from only one camera, the accuracy of the pose is less and also may be observed that the number of key points detected is less [Figure 6a].

It is observed that the accuracy of prediction using OpenPose was around 70%. OpenPose is preferred for 2D pose detection for multiperson system, which includes body, facial, foot, and hand key points.^[30] It is reported to have been used for vehicle detection as well. This method of pose estimation suffers estimating the poses when the ground truth has nontypical postures and also in estimating poses in crowded images, leading to the overlapping of key points. The number of key points detected is more than the EpipolarPose, yet during computation, the accuracy is compromised [Figure 7a] as Graphics processing units (GPU)-powered systems were not used.

It has been reported that after using the fully connecting layer to detect the features using PoseNet, the results have worsened as the network was likely to overfit to the training data. In our work, PoseNet methods gave an accuracy of about 80%. Figure 8 shows the key point detection by PoseNet.

However, MediaPipe has better accuracy as compared to EpipolarPose, OpenPose, and PoseNet. It may be observed



Figure 6: Key point detection by EpipolarPose. (a) Ardhachandrasana. (b) Tadasana. (c) Trikonasana. (d) Veerabhadrasana. (e) Vrukshasana



Figure 8: Key point detection by PoseNet. (a) Ardhachandrasana. (b) Tadasana. (c) Trikonasana. (d) Veerabhadrasana. (e) Vrukshasana

from Table 1 that the reason for less accuracy for other methods could also be due to pose estimation using a single camera.

The background light and contrast also have an influence on the accuracy values; it is clear that MediaPipe provides better results and can estimate postures more accurately than other methods, and hence, it is the most suitable technique for pose classification. It is also observed that the accuracy of a few postures in the MediaPipe is also less because the MediaPipe does not detect the neck key point.

The accuracy of each of these could be increased further with an increase in the training dataset; but nevertheless, it clearly illustrates the comparative study between different pose estimation methods.

The present study used four different deep learning architectures, i.e., EpipolarPose, OpenPose, PoseNet, and MediaPipe, which are suitable for pose estimation to evaluate yoga postures, and the results support the fact that MediaPipe has better accuracy compared to the other methods despite using a single camera.

Further research would be needed to expand this technique for other advanced postures for pose estimation and correction using the same methodology which involves simple tools with better accuracy to assist individuals practicing yoga postures as a self-evaluation as well as a biofeedback mechanism.



Figure 7: Key point detection by OpenPose. (a) Ardhachandrasana. (b) Tadasana. (c) Trikonasana. (d) Veerabhadrasana. (e) Vrukshasana

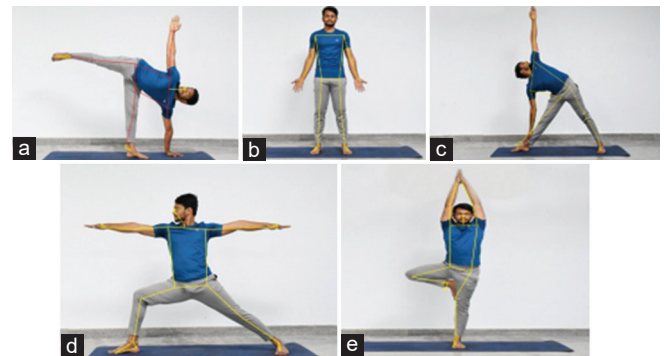


Figure 9: Key point detection by MediaPipe. (a) Ardhachandrasana. (b) Tadasana. (c) Trikonasana. (d) Veerabhadrasana. (e) Vrukshasana

Table 1: Comparison of postures with accuracy (%) of prediction

Postures	Accuracy of EpipolarPose	Accuracy of OpenPose	Accuracy of PoseNet	Accuracy of MediaPipe
Ardha Chandrasana	37.5	72.22	75	78.78
Tadasana	51.25	55.55	87.5	90.9
Trikonasana	58.75	66.66	81.25	85.75
Veerabhadrasana	62.5	72.22	81.25	81.81
Vrukshasana	56.25	77.77	87.5	88.81

Discussion

The present study used four different deep learning architectures, i.e., EpipolarPose, OpenPose, PoseNet, and MediaPipe, which are suitable for pose estimation to evaluate yoga postures, and the results support the fact that MediaPipe has better accuracy compared to the other methods despite using a single camera.

Muhammed *et al.*^[21] in their work used a self-supervised EpipolarPose pose estimation model which does not need 3D ground-truth data or camera parameters. During training, a 3D pose is obtained using the geometry of a 2D pose estimated from multiview images and used to train a 3D pose estimator. Furthermore, Yihui *et al.*^[32] proposed a differentiable epipolar transformation model where 2D is

detected to leverage 3D-aware features to improve 2D pose estimation.

Haque *et al.*^[33] used CNN to estimate the human pose present in a 2D image with an accuracy of 82.68, and Dushyant *et al.*^[34] reported on techniques using CNN to estimate 2D and 3D pose features using an architecture called SelecSLS Net and then predicted a skeletal model fit. Jose and Shailesh^[35] used 3D CNN architecture; a modified version of C3D architecture was used for pose estimation which gave an accuracy of 91.5%. Santosh Kumar *et al.*^[36] in their work reported an accuracy of 99.04% for using CNN for feature extraction and LSTM for temporal prediction.

Yoga is a form of physical exercise demands performing it accurately. Anilkumar *et al.*^[37] reported on a yoga monitoring system which is implemented to estimate and analyze the yoga posture where the user is notified of the error in the posture through a display screen or a wireless speaker. The inaccurate body pose of the user can be pointed out in real time, so that the user can rectify the mistakes. In this work, the nose is assumed to be the origin, so that all calculations are done with respect to the location of the nose in the image. An imaginary horizontal line passes through the nose's coordinates. This is the X-axis of all the angles and are calculated with respect to this horizontal line. However, in our work, we have divided the image into quadrants and compared the key points. Deepak and Anurag^[38] uploaded a photo of the user performing the pose and compared it with the pose of the expert, and the difference in angles of various body joints was calculated. Based on this difference of angles, feedback is provided to the user to improve the pose.

Chen *et al.*^[39] proposed a yoga posture recognition system using Microsoft kinetics to detect joints of the human body and to extract the skeleton and then calculated various angles to estimate the poses confirming accuracy of 96%. Chiddarwar *et al.*^[40] reported a technique for android application discussing the methodology used for yoga pose estimation. However, the present study demonstrated that MediaPipe has better accuracy compared to the other methods despite using a single camera.

Further research would be needed to expand this technique for other advanced postures for pose estimation and correction using the same methodology which involves simple tools with better accuracy to assist individuals practicing Yoga postures as a self-evaluation as well as a biofeedback mechanism.

Conclusions

The human pose estimation can be effectively used in the health and fitness sector. Pose estimation for fitness applications is particularly challenging due to the wide variety of possible poses with large degrees of freedom, occlusions as the body or other objects occlude limbs as seen from the camera, and a variety of appearances or outfits.

This work estimates the accuracy of different postures and compares them with four different architectures. Based on the results, the study concludes that the MediaPipe architecture provides the best estimation accuracy.

Acknowledgment

The authors would like to thank B N M Institute of Technology and SVYASA Deemed to be University for jointly collaborating toward the completion of this research work.

Ethical clearance

The study was approved by the Institutional Ethics Committee of Swami Vivekananda Yoga Anusandhana Samsthana (S-VYASA), Bengaluru (Approval Letter No: RES/IEC-SVYASA/193/2021.). The study procedure was explained and signed consent was obtained from the participants.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

- Collins C. Yoga: Intuition, preventive medicine, and treatment. *J Obstet Gynecol Neonatal Nurs* 1998;27:563-8.
- Nath A. Impact of COVID-19 Pandemic Lockdown on Mental Well-Being amongst Individuals in Society. *International Journal of Scientific Research in Network Security and Communication*. 2020;8.
- Loren F, Saltonstall E, Genis S. Understanding and preventing yoga injuries. *Int J Yoga Ther* 2009;19:47-53.
- Salmon P, Lush E, Jablonski M, Sephton SE. Yoga and mindfulness: Clinical aspects of an ancient mind/body practice. *Cognitive and behavioral practice*. 2009;16:59-72.
- Yash A, Yash S, Abhishek S. "Implementation of Machine Learning Technique for Identification of Yoga Poses," *IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*; 2020.
- Li L, Martin T, Xu X. A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. *Appl Ergon* 2020;87:103138.
- Cheung KM, Simon B, Takeo K. "Shape-From-Silhouette of Articulated Objects and Its Use for Human Body Kinematics Estimation and Motion Capture," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; 2003.
- Obdržálek Š, Kurillo G, Ofli F, Bajcsy R, Seto E, Jimison H, *et al.* "Accuracy and Robustness of Kinect Pose Estimation in the Context of Coaching of Elderly Population." *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; 2012. p. 1188-93.
- Georgios P, Axenopoulos A, Daras P. "Real-Time Skeleton-Tracking-Based Human Action Recognition Using Kinect Data," *International Conference on Multimedia Modeling*. Springer, Cham; 2014.
- Silvia Z, Freifeld O, Black MJ. "From Pictorial Structures to Deformable Structures," *IEEE Conference on Computer Vision and Pattern Recognition*; 2012.

11. Trumble M, Gilbert A, Hilton A, Collomosse J. "Deep Autoencoder for Combined Human Pose Estimation and Body Model Upscaling," Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 784-800.
12. Qize Y, Wu A, Zheng W. "Person Re-Identification by Contour Sketch Under Moderate Clothing Change," IEEE Transactions on Pattern Analysis and Machine Intelligence; 2019.
13. Zhao M, Li T, Abu Alsheikh M, Tian Y, Zhao H, Torralba A. Through-wall human pose estimation using radio signals. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p. 7356-65.
14. Li J, Wang C, Zhu H, Mao Y, Fang HS, Lu C. "Crowdpose: Efficient Crowded Scenes Pose Estimation and a New Benchmark," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p. 10863-2.
15. Sarafianos N, Boteanu B, Ionescu B, Kakadiaris IA. 3d human pose estimation: A review of the literature and analysis of covariates. Computer Vision and Image Understanding 2016;152:1-20.
16. Jiang Y, Li C. Convolutional neural networks for image-based high-throughput plant phenotyping: A review. Plant Phenomics 2020:Article ID 4152816. doi.org/10.34133/2020/4152816.
17. Manchen W, Tighe J, Modolo D. "Combining Detection and Tracking for Human Pose Estimation in Videos," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020.
18. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. Neurocomputing 2016 Apr 26;187: 27-48.
19. Liu L, Ouyang W, Wang X, Chen J, Liu X, Pietikäinen M. Deep learning for generic object detection: A survey. Int J Computer Vis 2020;128:261-318.
20. Muhammed K. "D/3D Human Pose Estimation Using Deep Convolutional Neural Nets," MS Thesis; 2019.
21. Muhammed K, Karagoz S, Akbas E. "Self-Supervised Learning of 3d Human Pose Using Multi-View Geometry," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019.
22. Chen CH, Tyagi A, Agrawal A, Drover D, Mv R, Stojanov S, *et al.* "Unsupervised 3d Pose Estimation with Geometric Self-Supervision," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p. 5714-24.
23. Hidalgo G, Raaj Y, Idrees H, Xiang D, Joo H, Simon T, *et al.* "Single-Network Whole-Body Pose Estimation," Proceedings of the IEEE/CVF International Conference on Computer Vision; 2019. p. 6982-91.
24. Luvizon D, David P, Tabia H. "Multi-Task Deep Learning for Real-Time 3D Human Pose Estimation and Action Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence; 2020.
25. Laga H. A survey on deep learning architectures for image-based depth reconstruction. arXiv 2019. preprint arXiv:1906.06113.
26. Clark RA, Mentiplay BF, Hough E, Pua YH. Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives. Gait Posture 2019;68:193-200.
27. Saini T. "Manoeuvring Drone (Tello ans Tello EDU) Using Body Poses or Gestures," MS Thesis, Universitat Politècnica de Catalunya; 2021.
28. Pauzi AS, Mohd Nazri FB, Sani S, Bataineh AM, Hisyam MN, Jaafar MH, *et al.* "Movement Estimation Using Mediapipe BlazePose," International Visual Informatics Conference. Springer, Cham; 2021. p. 562-71.
29. Xinghua C, Fuqiang Z, Xin C. Epipolar constraint of single-camera mirror binocular stereo vision systems. Opt Eng 2017;56:084103.
30. Pawang F. "OpenPose: Human Pose Estimation Method,". Available from: <https://www.geeksforgeeks.org>. [Last accessed on 2021 Nov 15].
31. Walch F, Leal-Taixé L. "Image-Based Localization Using LSTMs for Structured Feature Correlation," Conference: 2017 IEEE International Conference on Computer Vision (ICCV); October, 2017.
32. He Y, Yan R, Fragkiadaki K, Yu SI. "Epipolar Transformer for Multi-View Human Pose Estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; 2020. p. 1036-7.
33. Haque S, Rabby AK, Laboni MA, Neehal N, Hossain SA. "ExNET: Deep Neural Network for Exercise Pose Detection." International Conference on Recent Trends in Image Processing and Pattern Recognition. Springer, Singapore; 2018. p. 186-93.
34. Mehta D, Sotnychenko O, Mueller F, Xu W, Elgharib M, Fua P, *et al.* Xnect: Real-Time Multi-Person 3d Human Pose Estimation with a Single rgb Camera. arXiv preprint arXiv:1907.00837. 2019.
35. Jose J, Shailesh S. Yoga asana identification: A deep learning approach. IOP Conf Ser Mater Sci Eng 2021;1110:012002.
36. Yadav SK, Singh A, Gupta A, Raheja JL. Real-time Yoga recognition using deep learning. Neural Computing and Applications 2019;31:9349-61.
37. Anilkumar A, KT A, Sajan S, KA S. Pose estimated yoga monitoring system. 2021. Available at: SSRN 3882498. [Last accessed on 2021 Jul 08].
38. Anilkumar A, KT A, Sajan S, KA S. Yoga Pose Detection and Classification Using Deep Learning. Proceedings of the International Conference on IoT Based Control Networks and Intelligent Systems - ICICNIS 2021: LAP LAMBERT Academic Publishing; 2020.
39. Chen CH, Tyagi A, Agrawal A, Drover D, Mv R, Stojanov S, *et al.* "Unsupervised 3d Pose Estimation with Geometric Self-Supervision," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p. 5714-24.
40. Chiddarwar GG, Ranjane A, Chindhe M, Deodhar R, Gangamwar P. AI-based yoga pose estimation for android application. Int J Inn Scien Res Tech 2020;5:1070-3.

Supplementary Table 1: Summary of yoga interventions in Ashtanga yoga and general yoga arm

Practice	Duration
Asanas	
<i>Tiryaktadasana</i> (swaying palm tree pose)	3 min
<i>Trikonasana</i> (triangle pose)	3 min
<i>Konasana</i> (angle pose)	3 min
<i>Padahasthasana</i> (hand to foot pose)	1 min
<i>Ardhahalasana</i> (half plow pose)	3 min
<i>Padavrttasana</i> (cyclical leg pose)	6 min
<i>Dwicakriasana</i> (cycling pose) repetitive	3 min
<i>Markatasana</i> (monkey pose)	6 min
<i>Bhujangaasana</i> (cobra pose)	3 min
<i>Salabhasana</i> (locust pose)	1 min
<i>Chakkiasana</i> (mill churning pose)	3 min
<i>Sthitta konaasana</i> (static angle pose)	3 min
<i>Sthitta konaasana</i> (static angle pose)	1 min
<i>Paschimottanasana</i> (seated forward bend pose)	1 min
Pranyamas	
<i>Ujjayi</i> (victorious breathing)	3 min
<i>Anulom-vilom</i> (alternative nostril breathing)	6 min
<i>Brahmari</i> (humming breath)	3 min
Meditation	
<i>Nadanusandan</i> (A-U-M chanting)	30 min/once per week
Relaxation	
Guided relaxation technique	3 min

Asanas are repeated 5-10 times within the stipulated duration with holding time of 10-15 s. Apart from these, the AY arm received an Ashtanga yoga-based orientation (eight limbs of yoga) which included discussions on Yama, Niyama, Asana, Pranayama, Pratyahara, Dharana, Dhyana, and Samadhi. The investigators explained the role of each component of Ashtanga yoga in maintaining good health and live a meaningful life. AY=Ashtanga yoga

Supplementary Table 2: Overview of the points discussed in orientation program

Limbs of yoga	Superficial meaning	Points discussed
Yama	Moral disciplines	How to build self-discipline that will be beneficial to others around us and how that can help in calming/toning the mind
Niyama	Positive observances	Discussed on the duties towards one's self and how it will help in navigate in life
Asana	Postures	How to align the participants' postures and win over the body
Pranayama	Breathing techniques	How to achieve freedom over breath and regulate emotional breathing
Pratyahara	Withdrawal of senses	The importance of control over senses to achieve higher state of mind
Dharana	Focused concentration	How to utilize the first five limbs of yoga on building focus
Dhyana	Absorption	How to meditate on one-self and stay away from interruptions
Samadhi	Bliss	How to cultivate the habit of staying above likes, dislikes, hatred, love, and treat everything equally