Chapter 1

Introduction

1.0 INTRODUCTION

The practice of Yoga involves various techniques which takes an individual through a systematic approach of moving from a grosser to subtle dimension of experience. As part of physical activity Yoga includes mainly the postures (Asana), often practiced either in a dynamic or a static way. Simple stretching and joint loosening exercises are done as preparatory practices. Asanas are generally followed by relaxation techniques and sometimes accompanied by voluntarily regulated breathing (Pranayama), and meditation. Yoga in this form has become familiar across the world, especially in America and Europe.

The practice of Asanas has evolved systematically from the medieval *Haṭha Yoga* period, and is sometimes so named, but it is comprehensively called as "Yoga". Since the mid-20th century, Yoga has been used, especially in the Western world, as physical exercise for fitness and suppleness (Ivtzan & Jegatheeswaran, 2015). Since each Yogic posture can be boiled down to a specific orientation in which the person's limbs need to be in, a Pose estimating program can guide a person through the various stances of a yogic cycle. The physical movements in yoga should be accompanied with slow bodily movement, rhythmic breathing and inward awareness (Omkar et al., 2011).

Nowadays People in all age groups, young and old have some pressure and anxiety in their studies or in their workplace which could lead to depression. People try to deal with this situation by involving in sports or by entertainment activities, but it may not be an effective solution to deal with natural depression. Yoga is an effective workout practiced across the globe to enhance physical, mental, and spiritual wellbeing which connect body, mind and the soul. It is important to practice yoga subtly and calmly to reap its benefits to the fullest and not otherwise (Kirk, Boon & DiTuro, 2005). Improper breathing and practice of yoga due to lack of guidance or prior knowledge to perform the posture could push themselves beyond their

flexibility limits and are most likely to strain their muscles which could not only lead to pain but also has an adverse effect on body joints by injuring them. It is important to perform yoga in a sequential procedure under suitable guidance for better benefits which requires an instructor to monitor and guide a person while practicing it. This process involves hiring a yoga trainer to monitor the sessions for productive results which may be afforded by many. Computer vision techniques based artificial intelligence to monitor and guide a person to practice yoga properly by providing feedback for posture correction in real time is the need of the hour (Garbett et al., 2021). An Artificial intelligence-based yoga instructor could solve the purpose of practicing yoga at one's own place and convenience.

Human Body Modelling

Human body modelling is the capability of making a computer to mimic human appearance for practical applications. It is a model made by the computer which can be used as a proxy for human body involved in problems associated with object detection and representation. Processing the features of the human body are quite complex as it involves wide range of shapes varying across subjects. Making a computer to understand people and their movements in images and video is a challenging and complex task made easy by Computer Vision for human-computer interaction. Human body modelling is a method to map probabilistic body parts with nodes used in graphics and in human pose estimation and detection.

Modeling the human body is an important aspect in human pose which locates the different joints in the body skeleton from an image. Preferred method of pose estimation is by using kinematic models where the body's kinematic structure and shape information is represented by its joints and limbs (Cheung et al., 2003). Different ways in which human body modeling could be done are shown in Figure 1.

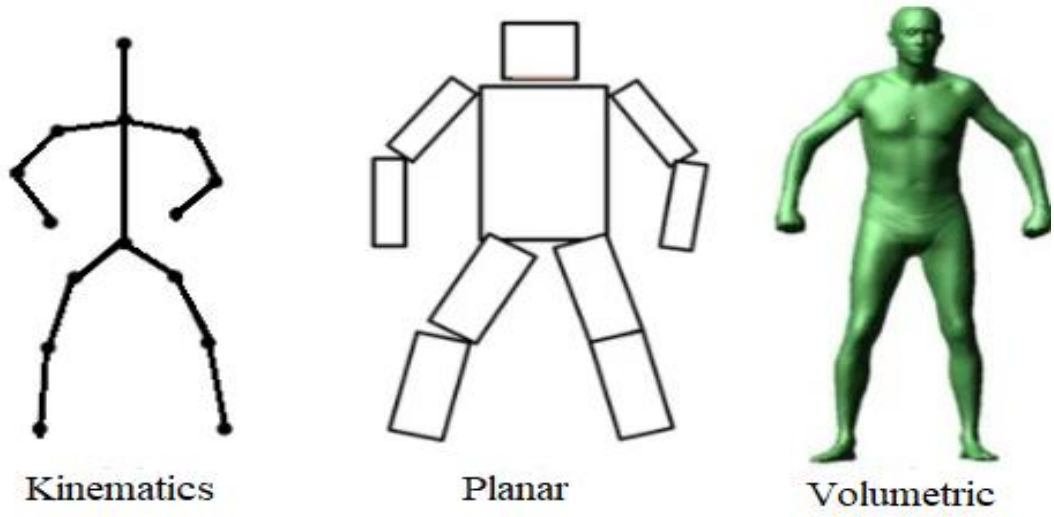

Figure 1: Human body modelling

The human body can be modelled 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 (Obdrzalek et al., 2012) (Papadopoulos et al., 2014).

However, 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 (Zuffi et al., 2012). The Volumetric Model represents a 3D model of well-articulated human body shapes and poses (Trumble et al., 2018). The challenges involved in human pose estimation is that the joint positions could change due to diverse forms of clothes, viewing angles, background contexts, and variations in lighting and weather (Yang et al., 2021) making it a challenge for image processing models to identify the joint coordinates and especially difficult to track small and scarcely visible joints.

Pose estimation (alignment and proficiency)

Pose Estimation is a technique used to identify the accurate performance of yoga postures performed by a person (Agrawal et al., 2020). Pose estimation is one of the long-standing

problems which have interested researchers over the last several decades. If we know the pose of a human, we can further infer relative positions of the limbs, elbow, wrists etc., and generate a pose model that can be used to perform smart surveillance. Human Pose Estimation is defined as the search for a specific pose in space of all articulated poses.

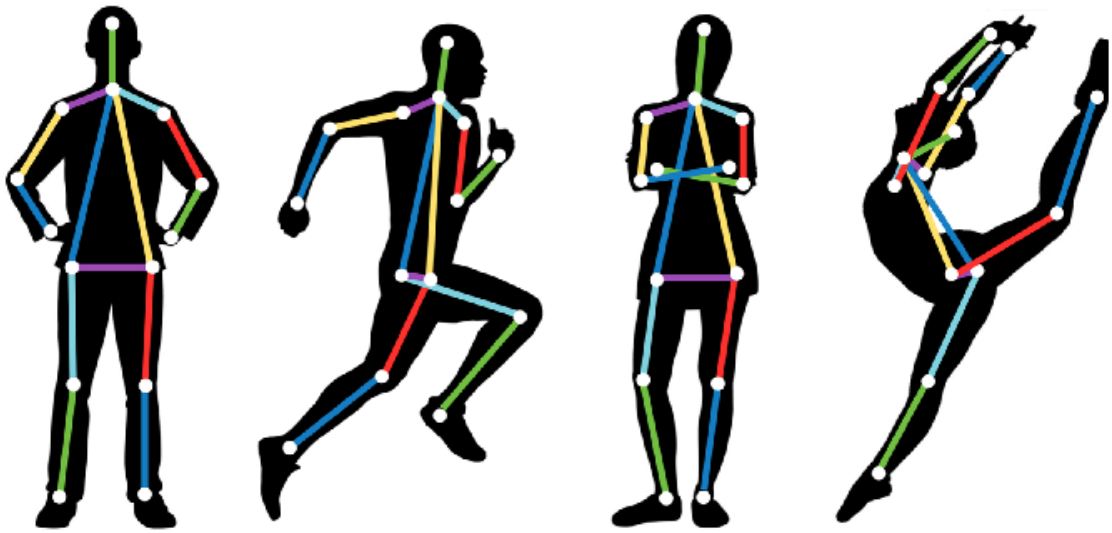

Figure 2: Examples of Human pose estimation taken from (*Https://Www.Analyticsvidhya.Com/Blog/2021/10/Human-Pose-Estimation-Using-Machine-Learning-in-Python/*, n.d.)

Figure 3: Example of key point extraction for single pose estimation taken from (*Https://Www.Analyticsvidhya.Com/Blog/2021/10/Human-Pose-Estimation-Using-Machine-Learning-in-Python/*, n.d.)

A Human Pose Skeleton represents the orientation of a person in a graphical format. It is a set of coordinates that can be connected to describe the pose of the person. Pose Estimation refers a computer vision techniques used to detect human figures in images and videos, so that one could determine, for example, where someone's left shoulder, right knee, elbows, wrist, etc shows up in an image (Toshev & Szegedy, 2014).

With the advancement in machine learning algorithms computer can now understand the body language by performing pose estimation and detection. Such advancements can significantly add value to applications in security, business intelligence, entertainment, sports and greatly in health and safety. One such area where such a technique could be used is in yoga posture detection in real time. Single pose or multi-pose estimation can be done, where a single object is estimated by the Single pose estimation method, and multiple objects by multi-pose estimation (Augusto et al., 2010). Human posture assessment can be done by mathematical estimation called generative strategies, also pictorially named discriminative strategies. Image processing techniques use AI-based models such as convolutional neural networks (CNNs) which can tailor the architecture suitable for human pose inference (Kishore, D. M., Bindu, S., & Manjunath, 2022).

How Does Human Pose Estimation Works?

Pose estimation is done by using fundamental algorithms which uses human pose and its orientation to predict. It involves 2 stages of which the first one is a framework to identify boundary boxes and then evaluates the pose in each box. The second stage involves estimating the key points for a person and then jointed, such as the elbow, knees, wrists, etc. Single or multi-pose can be estimate based on the application (Y. Chen et al., 2018). In single pose estimation, the model estimates the poses for a single person in a given scene as shown in Figure 3. In contrast, in the case of multi-pose estimation, the model estimates the poses for multiple people in the given input sequence as shown in Figure 4.

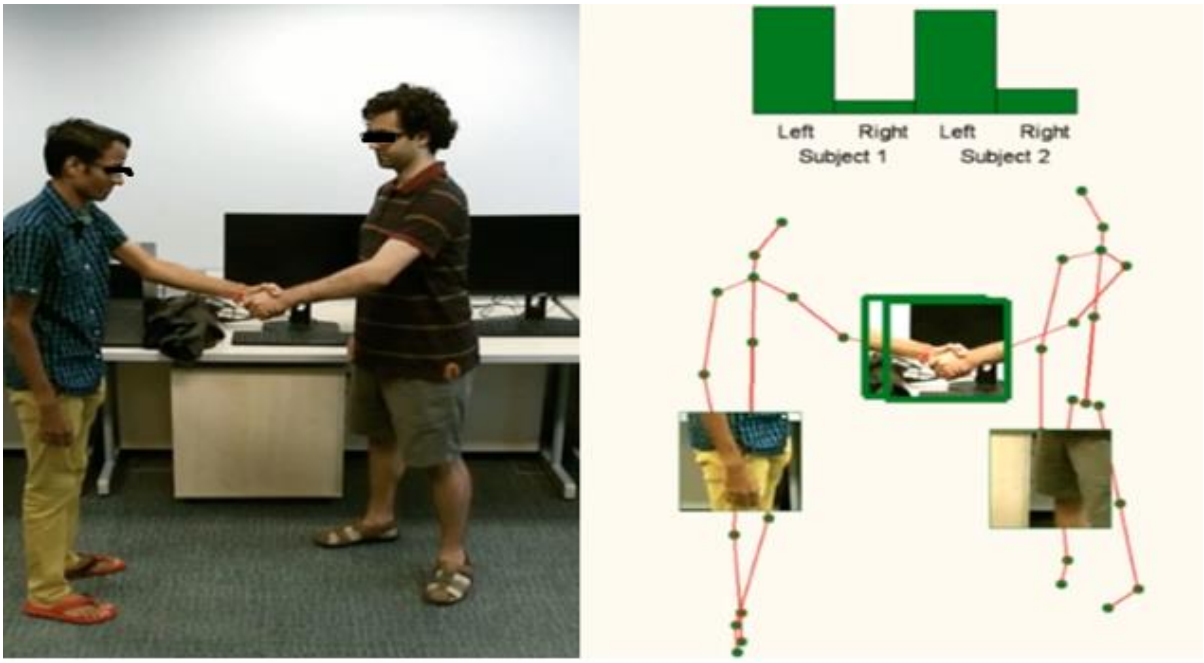

Figure 4: Example of key point extraction for Multi-pose estimation.

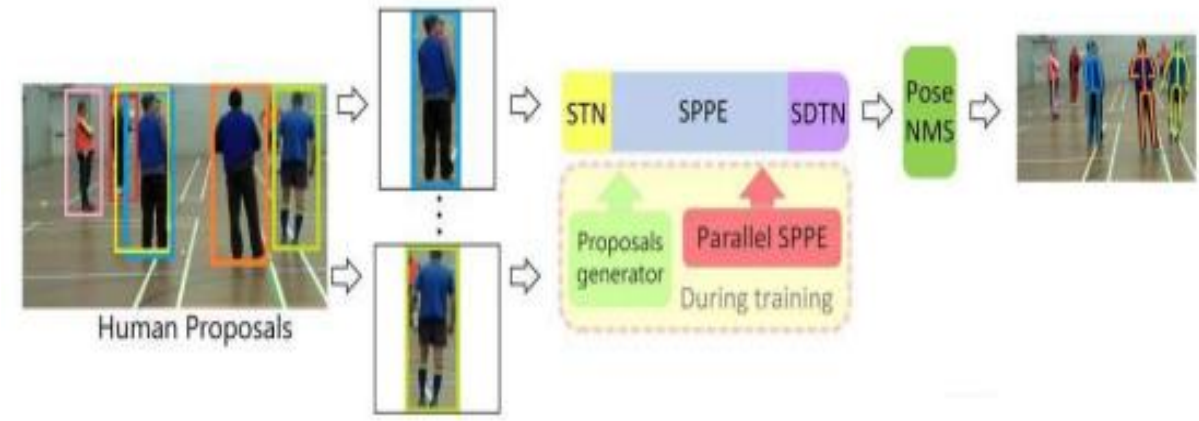

Figure 5: Top-Down approach Pose Estimation as taken from (Fang et al., 2017)

An approach for pose estimation can be done either by bottom-up or by top-down methods. In the Bottom-up approach, body joints are first estimated and then grouped to form unique poses, whereas, in top-down methods, first boundary box is detected and only then body joints are estimated (Fang et al., 2017).

In multi-pose estimation top down approach is commonly used where a separately trained

human detector is applied to detect each person and then run pose estimation on every detection. Top down approaches involve a segmentation step at the start, where each human is first segmented into a bounding box, followed by pose estimation being performed individually on each bounding box (Fieraru et al., 2018). Top down pose estimation can be classified into generative body-model based and deep learning-based approaches. Generative body-model based approach involves trying to fit a body-model on the image, allowing the final prediction to be human like (Chang et al., 2018).

The Figure 5 shows the flow of operations done on an image to display a final skeleton on a human. In contrast, bottom-up approach recognizes human poses from pixel-level image evidence directly. Bottom up approaches involve first detecting the parts or joints for one or more humans in the image, and then assembling the parts together and associating them with a human. In simpler terms, the algorithm first predicts all body parts/joints present in the image. This is typically followed by the formulation of a graph, based on the body model, which connects joints belonging to the same human.

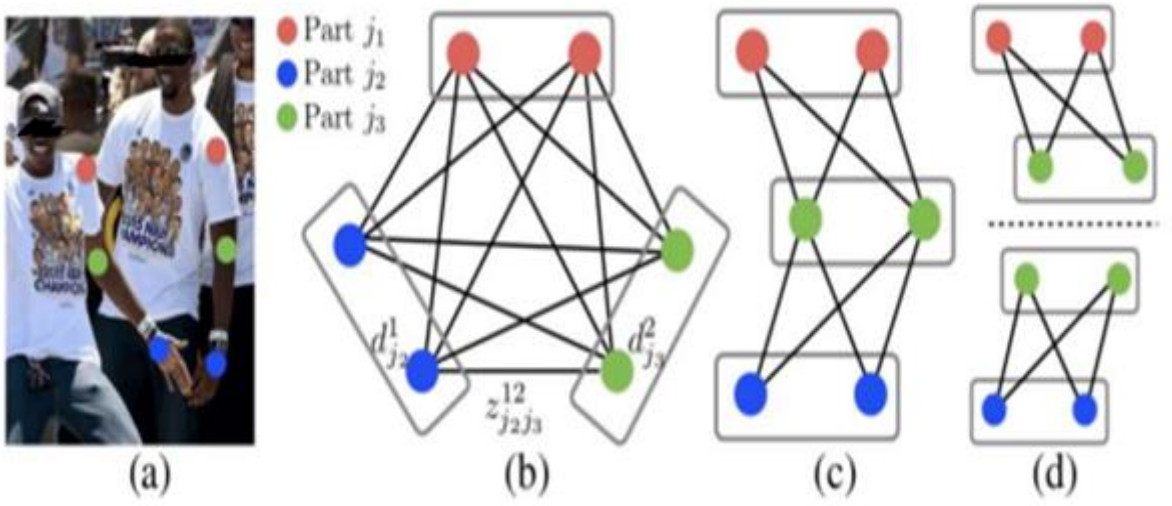

Figure 6: Bottom-Up approach Pose Estimation (Leunikau, A., Nedzved, A., Belotserkovsky, A., & Sholtanyuk, 2021)

The Figure 6 shows graph matching in a series of steps.

(a) shows the original image with the part detections (shown by coloured dots).

(b) shows the k-partite graph.

(c) shows the tree structure.

(d) Displays a set of bipartite graphs.

We use the Bottom-Up method to estimate or the position of human limbs and joints to create a program that can be implemented on any embedded device to teach a person to go through the positions of a yoga exercise using an "AI" driven instructor.

Pose Estimation with Deep Learning

Deep learning is a kind of machine learning technique used to teach the computer to learn different tasks directly from images, text, or sound. Deep learning methods are used to address tasks in computer vision, such as object detection, face recognition, action and activity recognition, and human pose estimation. Deep learning is an important element of predictive modeling which gets one more step closer with artificial intelligence to be used in applications which could not have been thought of earlier. For example, it is used in driverless cars to recognize different signs such as stop or to recognize pedestrian lamppost. With the help of huge dataset deep learning could be used in customer support, image processing, speech recognition, object recognition, natural language processing etc. Deep learning solutions have shown better performance in pose estimation (Guo et al., 2016) (Liu et al., 2020).

Deep learning methods use neural networks, designed to work the way human brain works. For example, the way human brain recognizes handwriting or facial recognition or recognize male or female and make quick decisions (Zhang et al., 2020). The word "deep" signifies the capability of hidden layers in the network used in models to train large datasets directly from the data without extracting the features. The popular neural network used in deep learning is convolutional neural networks (CNN) which learns from the input data accurately making

the architecture suitable to process images. CNNs use several hidden layers to learn different features from an image. It may learn edge detection in the first layer, some other parameter of the image in the second layer and finally learn some complex shapes of the object we are trying to recognize in the last layer. More the no of hidden layers better is the learned features (Jogin et al., 2018).

There are several deep learning-based pose detection methods developed for customised pose estimation. The existing reported architecture includes OpenPose, High-Resolution Net (HRNet), DeepCut, Regional Multi-Person Pose Estimation (AlphaPose), DeepPose, PoseNet, DensePose, TensorFlow, Epipolar pose and Mediapipe. The architecture for pose estimation could use bottom up or top down approaches which are featured to detect specific object in an image, or detect poses in spite of inaccurate bounding box, or real time multi-person pose estimation, or detecting both single and multi-person poses in images or video fields or on tensorflow.js to run on lightweight devices such as the browser or mobile device.

Pose estimating 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. These algorithms can be implemented on any embedded system, including smartphones, making it possible for commercial use in the form of an Application on the Android or iOS platforms. Posture estimation tasks are challenging as they require creating datasets from which real-time postures can be estimated. This research work estimates the five asanas such as: Ardhacandrāsana (Half-moon pose), Tāḍāsana (Mountain pose), Trikoṇāsana *(*Triangular pose), Vīrabhadrāsana (Warrior pose II) and Vṛkṣāsana *(*Tree pose) performed by the participant using four different deep learning architectures such as Epipolarpose, Openpose, Posenet, and Mediapipe. Deep learning architectures have been trained for above mentioned five asanas suitable for real-time and practical applications.

Pose estimation is done using OpenCv to detect the position of the person, using the detected position to create a skeleton that represents the “position” of the person by mapping their focal points. The mapping is done using model developed by Google and then using the skeleton as an image input to compare against the reference model trained to ascertain the degree of closeness of the position being performed to the ideal position. All of this is done using a model trained using Tensorflow.

Pose Correction

Pose correction deals with programming used to predict the actual poses performed by the user, in real-time, and provide corrective steps to do yoga accurately.

The technique uses for pose correction is shown by a flow chart as in Figure 7 which is self-explanatory. The graphical User interface takes a video of the user and passed to the mediapipe architecture for pose estimation. The essential key points at very joint of the body is determines and a skeleton is drawn. Using Euclidean distance, the angle is calculated and compared with the reference model. Cosine law of triangles are used to calculate the angles for reference and actual value. The angles at every joint are displays the results on the screen. Interactive pose correction is provided by using voice -to -text and text-to-voice model from Google.

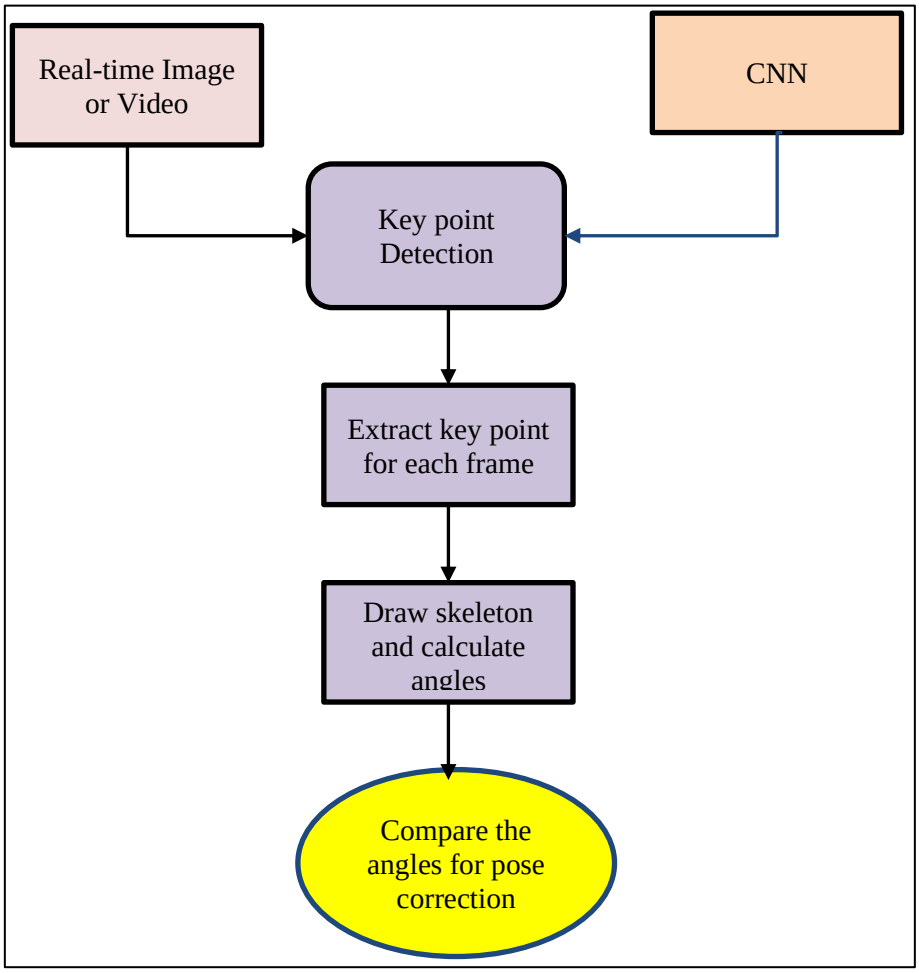

Figure 7: Figure showing pose correction technique

Summary

Yoga is increasing progressively over the last few, the number of highly skilled teachers has always been in scarcity leading to wrong ways of practicing yoga resulting in adverse effects. Attempts are being made to use technology to evaluate Yoga *Asana*as used in Biomechanics and Kinesiology which are predominantly adopted in sports medicine and fitness training. Such methods can also help individuals to understand their capacity to do such physical postures. Hence, there is a need for developing an *Äsana* evaluation tool which can help individuals to understand while doing their self-practice, help institutions to use while training teachers and therapists to understand the process of introducing yoga therapy. Several pose estimation architectures have been explored for poses estimation using artificial intelligence to emulate a Yoga Instructor/therapist who is capable of discerning if the position being performed by the person is correct or not.