# Yoga Pose Estimation using CNN and Mediapipe

**Shreyas Jain**
**Artificial Intelligence and machine learning**
**R.V. College Of Engineering**
**Bengaluru, Karnataka**
**shreyasjain.ai22@rvce.edu.in**

**Roopa Iranna Bagalkoti**
**Artificial Intelligence and machine learning**
**R.V. College Of Engineering**
**Bengaluru, Karnataka**
**roopaib.ai23@rvce.edu.in**

**K Preethi**
**Artificial Intelligence and machine learning**
**R.V. College Of Engineering**
**Bengaluru, Karnataka**
**kpreethi.ai23@rvce.edu.in**

*Abstract*—**Yoga pose detection holds significant importance in various aspects of the yoga practice and its integration with technology. The importance of yoga lies in its ability to promote physical health, mental well-being, stress reduction, improved focus, emotional balance, resilience, spiritual growth, and a holistic approach to life. With the increasing popularity of yoga, there is a growing need for technological advancements to support practitioners and instructors in monitoring and refining their practice. The paper begins by outlining the significance of automated yoga pose detection, highlighting the potential benefits it offers in providing real-time feedback, enhancing self-correction, and optimizing performance. It explores the existing literature on computer vision and machine learning techniques applied to human pose estimation and their applicability to yoga pose detection.**

***Keywords—yoga pose, asanas, machine learning, deep learning, human pose detection, yoga pose recognition, CNN,MediaPipe***

## Introduction

Yoga originated from India and has become popular worldwide because of the numerous physical and mental health benefits it provides. Mindful movement, controlled breathing, and meditation together have been promoting holistic wellbeing. n. Yoga pose detection is an emerging research area that pertains to the development of automatic systems that identify and track the pose accurately. Yoga pose detection is a significant technology that holds great promise for changing the practice, teaching, and analysis of yoga. Yoga pose detection systems, through computer vision and machine learning techniques, analyze images or videos of practitioners to give real-time feedback on the alignment of pose, posture, and movement. Several potential benefits can be expected from this technology: self-correction, remote instruction, tracking progress, injury prevention, and a customized yoga program. This research paper contributes not only to the practical implications for practitioners but also to the scientific understanding of yoga and its impact on physical and mental well-being. We will attempt to explore the underlying mechanisms through which yoga positions have a positive impact on the mind-body system by analyzing a wide range of empirical studies.

## Literature Survey

Human pose estimate research now encompasses a very important area of study known as yoga pose identification. Before the Deep Neural Network architecture and posture estimation frameworks were developed, attempts had been made to design automated and semi-automatic systems for the assessment of physical activity and sports.[11] Open pose is used to record the user and identify key points, CNN is used to identify patterns between key points in a single frame, and LSTM is used to memorise the pattern (Santosh Kumar Yadav: 2019). Deep neural networks have even transformed human pose estimation from the traditional systems. The deep neural network architectures are utilized in developing multiple posture estimation models, improving their performance and saving their cost. Open Pose is one of the most widely used open-source posture estimation models that emerged in (Cao et al.; 2017). (Santosh Kumar Yadav; 2019) utilized the Open Pose model for recognizing 15 critical points in yoga stance. Recovered key points were further used to classify an image to obtain a particular stance using the classifier hybrid model comprising CNN and LSTM. It utilizes the reference points to determine the pose angle and subsequently compute the error. [7] (Jothika Sunney:2022) uses Media pipe Blaze pose model outperforms existing pose estimation frameworks forreal-time Yoga/Fitness applications.

It outperforms open pose by factors of 25-75 with 20 cores in a single desktop CPU. Additionally, in

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

media pipe, the models tend to output three-dimensional keypoints. Therefore media pipe is available that offers 3D feature extractions without necessitating depth sensor hardware. Still, not enough research has integrated Media pipe together with Machine Learning algorithms. The novelty of this research lies in using a combination of [7] Media pipe Blaze pose model and [7] XgBoost classifier for the real-time yoga pose detection. In addition to that, this study evaluates how deep learning and machine learning models work on the 3D Landmark data that has been generated from Blaze pose model. The drawback of this model is that it cannot be used for multi-person detections. In the absence of enough publicly available datasets for yoga pose detection, research could have focused on creating a new dataset with more poses. As well, rather than passing key points directly to the model, this work can be improved by considering the angle between key points. Also, providing additional feedback on how to correct a yoga pose in real-time could improve real-time feedback. Currently, the Blaze pose model identifies only one person in an image frame. Further research should be conducted to detect multiple people. Using the media pipe library, the study first pinpointed important locations, then captured the coordinates in CSV format. The likelihood of each yoga stance for the current image sequence was calculated using the SoftMax level, and the pose with the highest probability was output. The model achieved high accuracy with a training accuracy of 99.49% and a testing accuracy of 99.70%. The study shows that the proposed model shows promise for real-time monitoring of yoga practices. Overall, the research shows that there is potential for further improvements in sentiment analysis through the use of larger data sets, additional numerical features, combinatorial methods, and advanced machine learning techniques such as deep and transfer learning.

Yoga poses: (a)Bhujangasana, (b)Padmasana, (c)Shavasana, (d)Tadasana, (e)Trikonasana, (f)Vrikshasana

The authors used Rectified Linear Unit activation functions and dropout regularisation to improve the network's performance. The authors trained the model using the Adam optimizer and a categorical cross-entropy loss function. The dataset was split into a training set (80%) and a validation set (20%) for the purposes of training and validating the model. The scientists also employed early stopping and model check points to avoid overfitting and to maintain the best performing model. The accuracy, precision, recall, and F1 score were only a few of the metrics the authors used to evaluate the YogaPoseNet model's performance. They also assessed the suggested model's performance in comparison to other deep learning models. [16] This paper highlights the significance of stance detection and recognition in yoga. Although there have been earlier attempts to identify yoga postures, the accuracy of these techniques has been constrained, according to the authors. They mention how other applications have showed promise for the use of CNNs for stance detection and speculate that it would work well for yoga pose recognition as well. As part of the methods outlined in the research, yoga positions are taught to a CNN. The scientists used data augmentation techniques to gather a collection of 10,000 images that represented 15 different yoga stances in order to increase the dataset's size. Then, training, testing, and validation sets were created from the dataset. Three convolutional layers, each followed by a max pooling layer and two fully connected layers, were used by the authors to train a CNN. They employed the SoftMax activation function for the output layer and the Rectified Linear Unit (ReLU) activation function (BhattacharyaS:2022). The categorical cross-entropy loss function and the Adam optimizer were used to train the CNN. Early halting was employed by the authors to avoid overfitting. The accuracy for the 15- class classification issue, as reported by the authors, was 91.7% after they evaluated the model's performance on the testing set.

# Methodology

An input for the suggested system is a real-time video series. Yoga posture predictions combined with advice on optimal angle and posture correction will be the outcome. Key point extraction, position prediction, and pose correction are the system's three core processes. Based on the user's position, the key point extraction phase seeks to identify and extract the locations of key points [9]. The phase of pose prediction categorizes whether or not the pose is valid and establishes the model architecture. stance correction is the last stage, during which the user is provided further feedback for correcting the stance and is shown the similarity percentage to the actual pose.

*A. Key points Extraction*

In the first stage, all of the video's frames are used to extract keyframes, which are then stored in CSV format. Examples of key points include shoulders, elbows, wrists, knees, and other body parts that are important for the creation of a yoga position. For the purpose of extracting key points, we used the cross-platform Media Pipe library. produced by Google that offers incredible pre-built machine learning (ML) solutions for computer vision challenges. In body posture estimation, the image is used to estimate the body configuration (pos). The following primary steps make up the condition evaluation process:

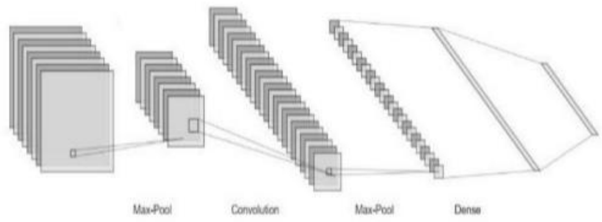
1) Identification of the major joints and points on the human body.

2) Classify these projected joints.

We can calculate the distance between key points (e.g., shoulder to elbow, elbow to wrist) using the Euclidean distance formula to measure the length of different body segments:

**Euclidean Distance: For two points A(x1, y1) and B(x2, y2): Distance = √ ((x2 - x1)^2 + (y2 - y1)^2)**

The different joints of the human body initially establish the essential points. It comprises all general positions such as the eye, ear, neck, and shoulder. In the second stage, by grouping all of these joints, the entire structure of the body should be organized, and these primary points of grouping should forecast the person's position at any given time.




Convolutional Neural Network

***Data Collection***

Media Pipe is used to extract key points, and the joint position values are saved in a CSV file. The CNN models are then utilized to forecast asanas. Combining results in the best mix of filtered CNN features.

***Pose Correction***

Following the classification of the projected posture as correct with respect to the specified pose, the user is given suitable feedback, and the similarity percentage (using cosine similarity) is calculated and reported to the user.

Cosine similarity is used to calculate the similarity between two vectors (e.g., the vector representing the user's pose and the vector representing the standard pose). It can be used to determine how closely the user's pose matches the correct pose:

**Cosine Similarity:**

**Cosine(Θ) = (A · B) / (||A|| * ||B||)**

**Where A and B are the vectors representing two poses, denotes the dot product, and ||A|| and ||B|| are the magnitudes of the vectors.**

Important and significant angles have been determined and rules have been developed for each of the six yoga positions included in the dataset, which will be discussed more below. A threshold is established for each rule, which represents the user's maximum deviation from the conventional stance. To establish the threshold for acceptable pose deviation, you can use a mathematical formula based on your specific requirements:

**Threshold = a * (Standard Deviation of Key Point Distances) 'a' is a scaling factor to control the threshold level.**

If the user reaches this threshold value, he or she receives feedback in the form of text and speech. Calculating the tangent inverse of the slope with positive X-axis yields the angle between two key points. Equation below gives the formula for calculating the angle between two key point coordinates:

**Angle (Θ) between two points (x1, y1) and (x2, y2): Θ = arctan((y2 - y1) / (x2 - x1)**)

The communication is initially received as text, which is then turned into speech using the Pyttsx3 package [41]; this is a text to speech converter that also works offline. The cosine similarity between the key points of the user's position and the standard position is determined in this work. As a result, this measurement represents the degree of similarity to the actual pose.

**Emperical analysis and result**

There are three key steps in this section. For the movies captured by the webcam in the first stage, frame-by-frame results utilizing the CNN model and -LSTM are displayed. The results of the second phase, 45, are predicted from recorded videos after images have been analyzed. Results for recorded videos are given following a query of 45 photos. Higher and more consistent findings are attained when 45 frames have been surveyed. Actual prediction outcomes are shown in the third stage. This model can predict six user posture sequences in a real-time setting and can validate the accuracy of the poses. The following factors are used to evaluate a yoga classification system:

A)**Classification Score:** The classification score typically represents the model's accuracy. It can be characterized as the proportion of accurate predictions to all input samples.

**Classification Score = (Number of Correct Predictions) / (Total Number of Input Samples)**

B) **Confusion matrix:** The correctness of the model is fully described by the confusion matrix. The confusion matrix can be used to determine metrics like as precision, recall, and precision score. and compute precision and recall using this data. Therefore, to score Fl, employ accuracy and recollection. It will be simpler to demonstrate how to calculate these things as a result. True negatives, true positives, false negatives, and false positives can all be counted. We always want 5 the diagonal of the matrix to have the highest number of samples since the diagonal values indicate correctly identified data.

**Precision = (True Positives) / (True Positives + False Positives)**

**Recall = (True Positives) / (True Positives + False Negatives)**

**F1-Score = 2 * (Precision * Recall) / (Precision + Recall)**

## CONCLUSION

The evaluation of the human condition has received a lot of attention lately. In contrast to other computer vision problems, human pose estimation requires that body pieces be constructed and placed according to a predetermined human body form. Assessing your posture while exercising can help you avoid injuries and perform better. We contend that a self-study yoga programme may both popularize and guarantee proper yoga practice. Deep learning techniques have potential as a result of substantial research in the area. An efficient real-time yoga monitoring system was presented in this paper. The Media Pipe library is used to first locate user-specific key points, then key coordinates are then captured and saved in CSV format. Then, we transmit a 45-image sequence that was created in real time to the model. The model in the example below uses CNN and LSTM in conjunction to identify useful characteristics and track the appearance of image sequences. The SoftMax level at the conclusion of calculates each yoga pose's probability for the current image sequence and outputs the pose with the highest probability of. More specifically, the user is given the output, which is 45 frames for each frame in counting mode. If the pose is deemed accurate, the user receives additional feedback based on a predetermined threshold. The threshold, in particular, is established to guarantee that the user maintains exact locations and angles while not overtaxing the machine. Finally, the user is shown a similar percentage of when compared to the default stance.

## FUTURE WORK

Only six yoga asanas are currently classified using the proposed models. There are many different yoga asanas, hence it is challenging to develop a posture assessment model that works for all asanas. By including more yoga positions done by people both indoors and outside, the dataset can be increased. The accuracy of the Open Pose pose estimation, which may not be effective when persons or body parts overlap, determines how well the models perform. This system can be implemented as a portable gadget that self-learns and makes predictions in real time. Performance detection for real-world applications is shown in this paper. Position can be determined using a similar method in tasks like sports, surveillance, healthcare, etc.

## REFERENCES


[1] Liaqat,S.,Dashtipour,K.,Arshad,K.,Assaleh,K.,& Ramzan,N.(2021,April 1) A Hybrid Posture Detection Framework: Integrating Machine Learning and Deep Learning IEEE

[2] Rajendran,A.K.,&Sethuraman,S.C.(2023 Februrary).A Survey on Yogic Posture Recognition.IEEE

[2] Swain,D.,Satapathy,S.,AcharayaB.,Shukla M.,Gerogiannis,V.C.,Kanavos &Giakovis,D.(2022,October31).Deep Learning Model for Yoga Pose Monitoring Algorithm

[3] Swain ,D.,Satapathy,S.,Patro,P.(2022,June 27).Yoga Pose Monitoring System using Deep Learning ,ResearchSquare

[4] Bhoite,P.,Kalekar,O.,Bhandari,S.,&Dhavas,N.(2023,April4 ).Yoga Posture Detection and Correction System International Reserch Journal of Modern Education and Technology

[5] Upadhyay, A., Basha, N. K., & Ananthakrishnan, B. (2023, February 17). Deep Learning-Based Yoga Posture Recognition Using the Y_PN-MSSD Model for Yoga Practitioners. Healthcare

[6] Kishore, D. M., Bindu, S., & Manjunath, N. K. (2022, September 5). Estimation of Yoga Postures Using Machine Learning Techniques. International Journal of Yoga

[7] Sunney, J. (2022). Real-Time Yoga Pose Detection using Machine Learning Algorithm.

[8] Bembde, M., Barude, S., Shinde, P., Thorat, T., & Thakar, D. (2022, July). Yoga Posture Detection and Correction System. International Journal of Advanced Research in Science, Communication and Technology, "www.ijarsct.co.in"

[9] Long, C., Jo, E., & Nam, Y. (2021, September 20). Development of a yoga posture coaching system using an interactive display based on transfer learning

[10] Narayanan, S. S., Misra, D. K., Arora, K., & Rai, H. (2021, May 13). Yoga Pose Detection Using Deep Learning Techniques. SSRN.

[11] Yadav, S. K., Singh, A., Gupta, A., & Raheja, J. L. (2019, May 9). Real-time Yoga recognition using deep learning. Neural Computing and Applications

[12] Kothari, S. (2020, May). Yoga Pose Classification Using Deep Learning (Master's thesis).

[13] D. Mohan Kishore, S. Bindu, Nandi Krishnamurthy Manjunath. (2022). "Estimation of Yoga Postures Using Machine Learning

[14] Kumar, D. (2020). "Yoga Pose Detection and Classification Using Deep Learning." In Proceedings of the IEEE International Conference on Artificial Intelligence and Machine Learning

[15] Murthy, N. A., Kumar, B. R. M., & Singru, N. M. (2015). "Real-time yoga posture recognition and correction system using Kinect sensor." International Journal of Computer Science and Information Technologies,

[16] Bhattacharya, S., Borkar, J. H., & Dnyaneshwar. (2021). "Yoga Pose Detection using Convolutional Neural Networks." Proceedings of the International Conference on Computer Vision and Pattern Recognition