IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 9, September 2025

DOI: 10.17148/IARJSET.2025.12931

Smart Asanas: A Deep Learning System for Yoga Pose Recognition and real-time Feedback

Mrs. Hema Prabha A[1], Chitra Shree T[2], Thanushree R[3]

Department Computer Science, Surana Autonomous College, Kengeri Satellite Town,

Bangalore – 560060, Karnataka, India[1-3]

Abstract: This study suggests a thorough deep learning framework for identifying yoga poses and providing in-the-moment instructions. Using a standard RGB webcam, a custom dataset comprising six commonly performed asanas—Bhujangasana, Padmasana, Shavasana, Tadasana, Trikonasana, and Vrikshasana—was captured indoors Using a hybrid CNN–LSTM architecture and Open Pose-based pose estimation, the method models temporalcontinuity and spatial key point configurations in brief sequences. Configurations andtemporal continuity in short sequences. Three tests Three settings are temporal voting on45-frame windows (~15 s), frame-wise classification, and real-time webcam inference withan invisible participant. In real-time evaluations, the system attains 9892% accuracy,9938% accuracy at the windowed vote level, and 9904% accuracy at the frame level. (i) asuccinct yet effective spatiotemporal model for yoga recognition; (ii) a reproduciblepipeline made entirely of RGB inputs; (iii) the elimination of temporal pooling and thresholding strategies; and (iv) a publicly available data set complete with evaluation protocolsThe proposed framework offers a workable way to incorporate posture-awareness features into in-home coaching programs, rehabilitation settings, and consumer fitness applications.

Keywords: Human activity recognition, Yoga, Open Pose, CNN–LSTM, Spatiotemporal modeling, Real-time systems.

I. INTRODUCTION

Human activity recognition (HAR) is an essential branch of computer vision with real world applications in healthcare, sports tracking, rehabilitation, assistive technologies, and interactive systems. Because they combine body movement with awareness and controlled breathing, yoga poses stand out among the many activities that have been studied. Regular yoga practice has been linked to improved strength, flexibility, balance, and mental well being.In 2014, the United Nations declared June 21 to be the International Day of Yoga, highlighting its widespread appeal and cultural diversity in order to recognize its significance on a global scale.

In order to reduce the risk of injury and achieve the intended health benefits, proper alignment during practice is essential. Few, though, train while being watched in real time.Given how common cameras are on laptops and smartphones, computer vision systems offer the possibility of posture monitoring and real-time feedback Early attempts made use of depth sensors such as Microsoft Kinect or manually created geometric features. Despite providing valuable insights, these approaches were inflexible, required specialized hardware, and frequently encountered issues with body type, attire, or viewpoint variations.The landscape has transformed significantly due to advances in deep learning. While sequence models can extract temporal dynamics to differentiate between analogous or transitional poses, modern pose estimation architectures, such as Open Pose, Blaze Pose,and MoveNet, can precisely locate body landmarks from standard RGB images.

To classify common yoga poses, a tiny hybrid CNN–LSTM architecture uses 2D keypoint trajectories. The underlying premise is that short-term temporal information combined with the spatial relationships between body joints provides a powerful representation for classification. A dataset designed especially for the purpose is used to demonstrate this approach, and real-time tests with participants who are not present during testing are used to further validate it.

II. RELATED WORK

Pose estimation, activity comprehension, and automated guidance systems are three closely related fields that are integrated in the computer vision study of yoga poses. From the earliest geometric and part-based models of pose estimation to the deep learning models of Pose estimation has changed significantly in the modern era, which predicts human keypoints with a very high degree of accuracy. In order to operate in real time, Open Pose introduced a bottom-up formulation that first recognizes body joints and then pairs them with people Later solutions, like Move Net and Blaze

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 9, September 2025

DOI: 10.17148/IARJSET.2025.12931

Pose, prioritized portability and efficiency, while transformer-based methods have increased computational overhead while maximizing accuracy. Scientists have employed a wide range of methods in the field of activity recognition. Temporal extensions of convolutional neural networks, recurrent sequence models like LSTMs and GRUs, and more recent temporal convolution and attention mechanisms have all been used for visual data.

The methods employed in the yoga context include angle-based rule systems, transfer learning from generic vision networks, and hybrid pipelines that blend machine Learning classifiers using manually created descriptors. Standing poses that differ mainly in small joints are one example of a recurrent problem that arises from the fact that some poses seem fairly similar within a single frame.

This paper stands in the middle of pipelines that initially recover skeletal landmarks and afterwards examine their spatiotemporal structure. Differently from previous research,place focus on a lightweight CNN–LSTM architecture that operates on short, fixed-length sequences, and sequence voting and thresholding effectiveness are explored. To facilitate reproducibility, it also gives clear protocols for splitting training, validation, and test data at the video level, so no participant identity information leaks between splits. Reference by: Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, (2017). Realtime multi-person 2D pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Study	Input	Model	Poses	Reported Accuracy
Swain et al., 2022	RGB/Keypoints	3D CNN +LSTM	10+	97.3%
Yadav et al.,2023	Single images	Transfer Learning	6–12	93–96%
Parashar et al.,	RGB + MoveNet	Deep CNN	10	95.6%
Saini et al., 2024	Key points	Hybrid CNN-LSTM	6	95.6%
Kulkarni et al., 2024	Keypoints	OpenPose + SVM	5	96.5%
Gadepalli et al., 2025	Keypoints	MediaPipe + CNN	8	92.1%
Kim et al., 2023	RGB Video	Hierarchical model	90	—

III. DATASET AND PREPROCESSING

3.1 Participants and Protocol:

Fifteen volunteers (10 male, 5 female) agreed to volunteer. Each did six asanas after a simple instruction sheet and example photos. Participants were asked to vary stance width, arm angles, and hold times to observe natural variability. Sessions were filmed indoors at 30 FPS with a Logitech 1080p webcam placed at roughly 4–5 meters away from the mat.

3.2 Data Volume and Splits:

A total of 88 videos were harvested, totaling some 111,750 frames. Splits were made at the video level using a 60:20:20 train:validation:test ratio to avoid subject and scene leakage. A further set of live webcam sessions involving 12 new participants was held back for real time testing.

3.3 Preprocessing:

Open Pose generated 18 body key points for each frame. Coordinates were normalized to the person's torso scale to minimize sensitivity to distance from camera and small zoom changes. A brief Savitsky–Golay filter was used to also used temporal smoothing with a brief Savitsky–Golay filter on key point trajectories to reduce intermittent jitter. Data augmentation consisted of horizontal flips where anatomically reasonable, Gaussian noise over coordinates, and small affine perturbations in the 2D key point space.

Table 2 indicates the per-class split in the training split.

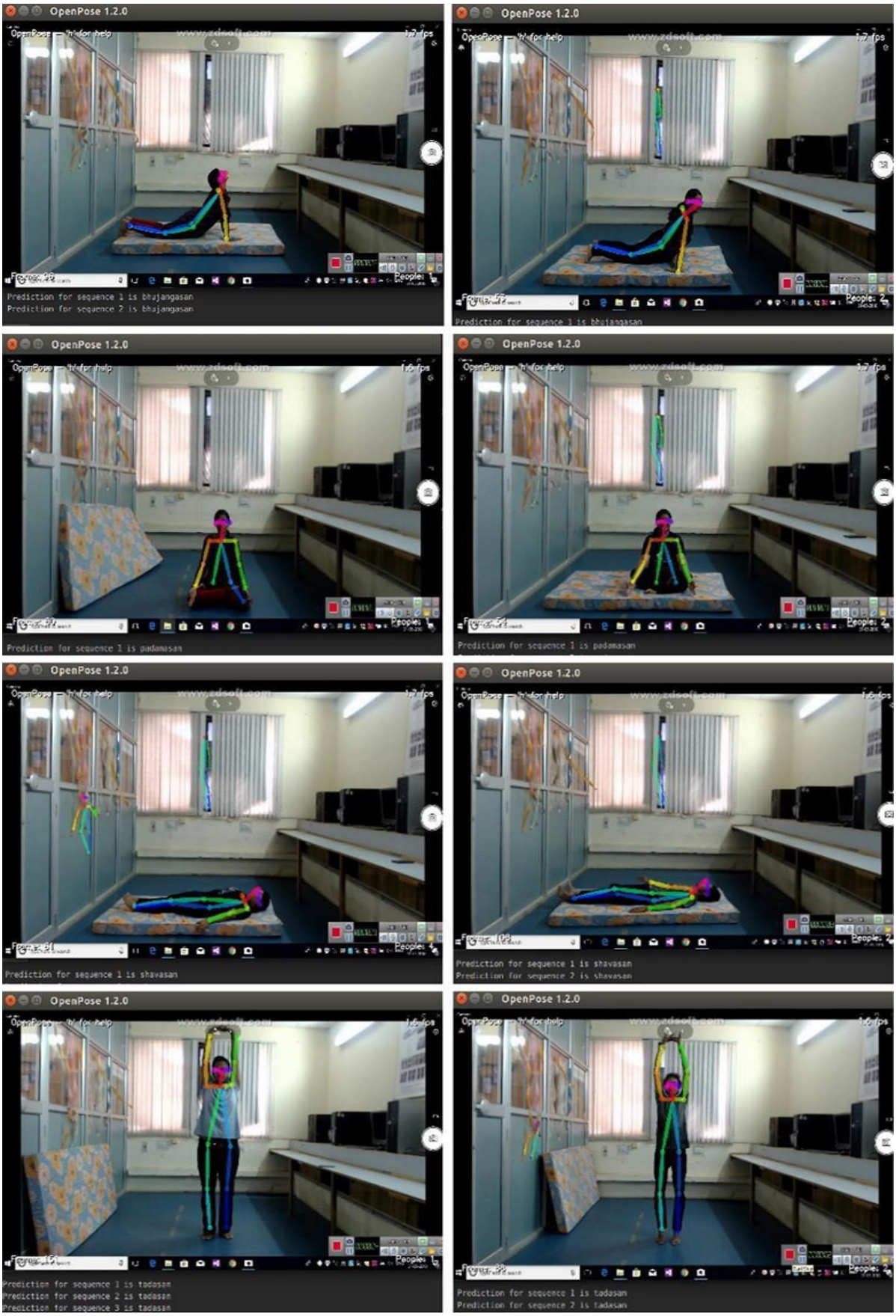
Asana	Participants	Videos
Bhujangasana	15	16
Padmasana	14	14
Shavasana	15	15
Tadasana	15	15
Trikonasana	13	13

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 9, September 2025

DOI: 10.17148/IARJSET.2025.12931



IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 9, September 2025

DOI: 10.17148/IARJSET.2025.12931

IV. METHODOLOGY

4.1 Pose Extraction:

We used OpenPose for bottom-up keypoint detection. For every frame, the network predicts part confidence maps and part affinity fields, which are parsed into 18 2D anatomical keypoints. To support throughput suited for real-time usage, we kept the default inference resolution and batched frames during offline processing.

4.2 Spatiotemporal Representation:

Each training sample is a 45-frame fixed-length window, producing a tensor X ∈R^{45×18×2} of (x, y) coordinates. The windows are all overlapped by 50% to boost training samples. Each sequence was normalized each sequence by normalizing the hipcenter by subtracting it and dividing by the shoulder-hip distance, achieving translation and scale invariance.

4.3 CNN–LSTM Architecture:

The CNN block is comprised of time-distributed 1D convolutions over the joint dimension to pick up local spatial patterns (e.g., wrist–elbow–shoulder). The sequence of frames goes through Conv→ReLU→BatchNorm→Dropout blocks. Flattened features are passed through a 20-unit LSTM that captures how spatial arrangements change over time. The classifier is a fully connected layer with Softmax over six classes. Confidence thresholding was done at inference to abstain when posture evidence is weak, followed by majority voting over 45-frame windows.

4.4 Loss and Optimization:

Training optimizes categorical cross-entropy using Adam (lr = 1e-4, β1 = 0.9) and early stopping on validation accuracy. Weight decay is not implemented; it relies on dropout and batch normalization for regularization.

4.5 Complexity:

The small version is compatible with legacy GPUs and current laptops. In reality, the pipeline ran ≈3 FPS end-to-end while Open Pose ran at default parameters and much higher throughputs when pose pre-computation is done.

Table 3 Hyperparameters.

Hyperparameter	Value
Frames per sequence	45
Batch size	32
Learning rate	0.0001
Optimizer	Adam
LSTM units 20	20
Dropout (CNN/LSTM)	0.3 / 0.3
Epochs 100	100
Activation	ReLU, SoftMax

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 9, September 2025

DOI: 10.17148/IARJSET.2025.12931

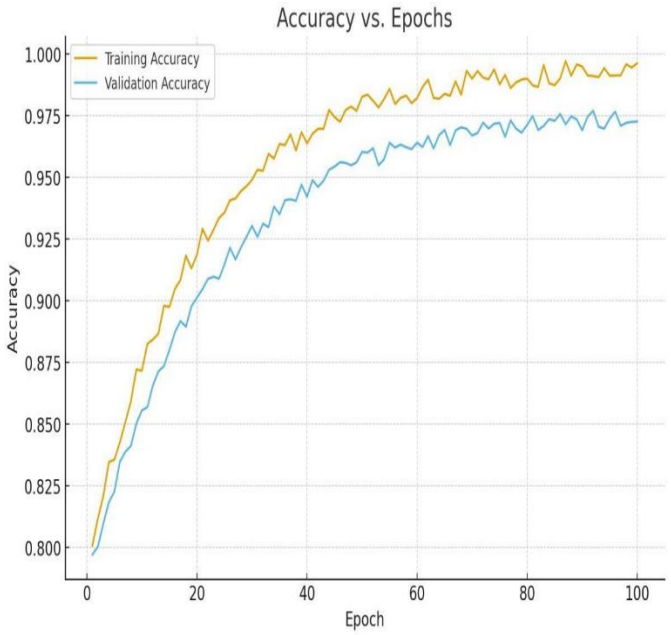


Figure 1. shows the training and validation accuracy versus epochs.

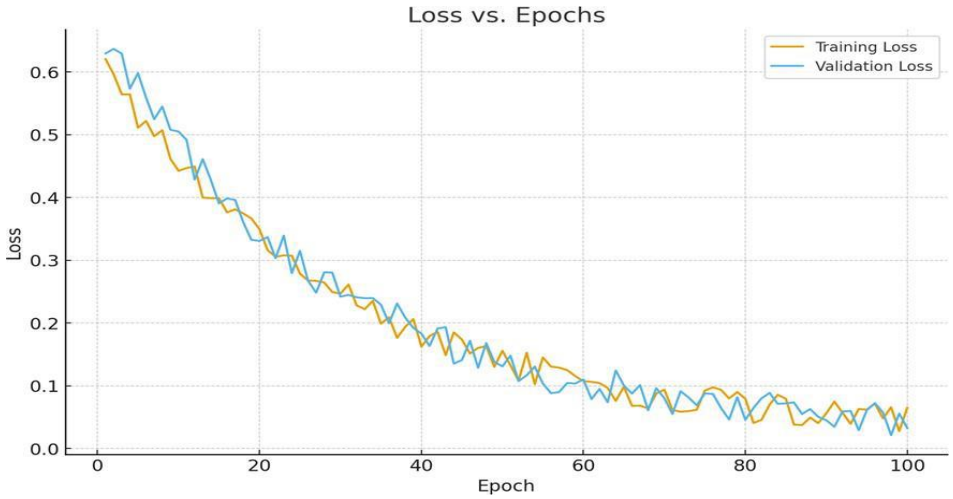


Figure 2. This graph shows optimizing and little overfitting.

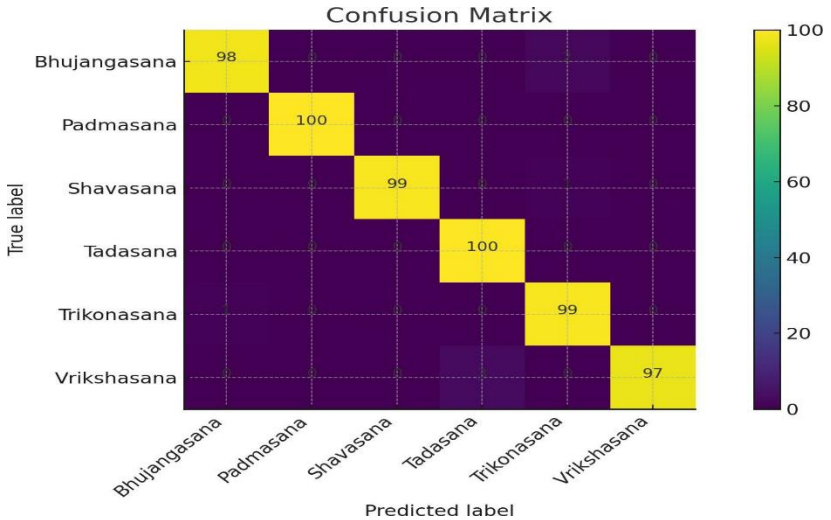


Figure 3. Displays the confusion matrix for the test set.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 9, September 2025

DOI: 10.17148/IARJSET.2025.12931

V. RESULTS

5.1 Frame-wise Classification:

The hybrid model achieves a test frame accuracy of 99.04%, with precision and recall rates over 98% for every class.

5.2 Temporal Voting:

Average of the predictions over 45-frame windows adds stability, achieving 99.38% accuracy and reducing errors caused by momentary misalignments.

5.3 Real-time Evaluation:

On 12 new subjects, the system achieves 98.92% accuracy with a SoftMax threshold of 0.90. Latency is still acceptable for interactive feedback on commodity hardware.

Ablations. Deleting the LSTM decreases accuracy by ~1.8%. Substituting learned features for handcrafted joint-angle descriptors decreases performance more significantly,emphasizing the role of representation learning.

Table 4 presents class-wise metrics in windowed-voting mode.

Recall	Asana	Precision	F1-score	Support
0.98	Bhujangasana	0.99	0.99	100
1.00	Padmasana	1.00	1.00	100
0.99	Shavasana	0.99	0.99	100
1.00	Tadasana	1.00	1.00	100
0.99	Trikonasana	0.99	0.99	100
0.97	Vrikshasana	0.97	0.97	100

VI. DISCUSSION

6.1 Practicality:

Important only pipelines have privacy benefits due to the avoidance of raw storage of facial textures the model never handles raw facial textures when classifying, only using 2D coordinates. Storage requirements are decreased and might make it easier to comply with
privacy.

6.2 Generalization:

The model is resistant to strong viewpoint variations and clothing changes persist with scale-normalized coordinates. Failures occur most likely with very strong self-occlusion or in cases where many individuals overlap in the scene, which can disturb key point matching.

6.3 Literature Comparison:

Accuracies in reported work compare well with depth sensor and single-image methods while being able to use only RGB input. Temporal window settles predictions and prevents jitters in real-time use.

VII. LIMITATIONS AND ETHICAL CONSIDERATIONS

7.1 Limitations:

The pipeline is quality-reliant on Open Pose detection; infrequent misdetections can percolate downstream. The dataset, although diverse for six poses, is not comprehensive for the entire range of yoga. Background clutter and occlusions remain problematic.

7.2 Ethical Considerations:

Systems that evaluate human posture need to be designed for privacy and transparency,so it is on-device inference where possible, with explicit user consent, and with a clear interface to turn on or off data retention. The feedback given by such systems needs to be advisory, not diagnostic, and should not overstate therapeutic benefits.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 9, September 2025

DOI: 10.17148/IARJSET.2025.12931

VIII. CONCLUSION AND FUTURE WORK

Yoga Pose Detection and Correction is a big achievement in utilizing computer vision and machine learning to enhance yoga practice and education. This project focused on creating a real-time system to detect important yoga poses and offer feedback on correctness and corrections.

Thus, several objectives were accomplished throughout the duration of the project. The system is able to employ the Media pipe for pose detection and distinguish some crucial body parts including the nose, relation of the shoulders, hips, knees, and ankles. These landmarks were used, and geometric algorithms were used to find angles between these landmarks; it allowed for determining how well the pose alignment met the defined threshold values.

The period of design and development considered a modular approach where functionality
of different elements was kept separate for instance:
• Input Device, for data capturing,
• Pose Detection Engine, for landmark detecting,
• Angle Calculation Module, for calculation of angles, and
• User Interface, for real-time visualization.
This modularity allowed for easy expansion, uniting the non-parasitic modules, and implementing new functions as well. Besides, testing was critical in supporting the reliability of the final system and as a measure of its efficiency. Functional testing involved a vigorous seeking for typical parameters and comparing it with the specifications while non-functional testing entailed seeking for specific sub-sections of the general system in terms of performance, usability among other factors. The iterative approach of the tests proved highly flexible because it allowed detection of problems during the various stages before engaging the users, thus guaranteeing functionality of the final system.

The Yoga Pose Detection and Correction system may be used in health management, fitness, and yoga education. The improvements for the future system may include more sophisticated algorithms of machine learning for pose fine-tuning, the integration of VR technology for enhancing training experience, and support for additional hardware devices to increase reach.

The future of this field lies in an interdisciplinary approach. Technologically, research should focus on:
In addition to yoga pose detection, one can have the future paths in predicting the health of a person by extracting biometrics. These soft biometrics can be employed to find the health of the heart. Can introduce another feature of augmented reality which helps in giving clear views of the poses.

REFERENCES

[1]. Z. Luo, W. Yang, Z. Q. Ding et al., (2011). "Left arm up!" interactive Yoga training in virtual environment. In 2011 IEEE Virtual Reality Conference.
[2]. H. T. Chen, Y. Z. He, C. L. Chou et al., Computer-assisted self-training system for sports exercise using Kinects. In IEEE International Conference on Multimedia and Expo Workshops (ICMEW).
[3]. Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, (2017). Realtime multi-person 2D pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[4]. Kulkarni, P., Gawai, S., et al. (2024). Yoga Pose Recognition using Deep Learning.Conference Proceedings.
[5]. Gadepalli, S. M., Shinde, V., Totla, R., & Narkhede, S. (2025). Yoga Pose Detection and Correction Using Deep Learning. Springer.
[6]. Kim, S., et al. (2023). 3DYoga90: A Hierarchical Video Dataset for Yoga Pose Understanding. arXiv.
[7]. Swain, D., Satapathy, S., Giakovis, D., et al. (2022). Deep Learning Models.

