WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Comprehensive Review of Real-Time Yoga Posture Correction Using Pose Detection and CNN–LSTM Models

Mohd Salauddin[a], Sanjay Singh[b], Sneha[c], Dr. Pardeep Chouksey[d], Praveen Sadotra[e], Mayank Chopra[f]

[a-f] Department of Computer Science & Informatics, Central University of Himachal Pradesh, Kangra, Himachal Pradesh, India

ABSTRACT :

AI and ML have revolutionized RTYPD and corrective feedback systems by enabling fast evolution in real-time.AI and ML have helped to revolutionize RTYPD and corrective feedback systems, enabling rapid evolution in real time.The focus of this review is on the studies published since 2019 up to 2025. When dis-cussing recognizing yoga poses, artificial intelligence plays a significant role, and making real-time inferences, and even fixing mistakes are at the heart of this technique. This document critically reviews 15 peer-reviewed studies on the topic of yoga pose recognition and surveys the variety of approaches that have been under-taken so far such as the Keypoint-based estimation of poses, OpenPose, MediaPipe, PoseNet and MoveNet, deep-learning models using CNNs and CNN-LSTM hybrids, 3D-CNNs and two-stream architectures, and the angle-based correction frameworks. The findings indicate that deep learning models that are coupled with lightweight pose detectors can achieve high accuracy and real-time execution of mobile and edge devices. However, there are still many areas that need to be improved, such as the range of datasets being signifi-cantly imbalanced, generalization, inter-body robustness and corrective feedback reliability. Moreover, the majority of the current systems are based on the identification of postures and not the actual biomechanical correction. This is a review of the current methods and a comparison of performance trends. It also find out research gaps and highlights future directions. They range from multimodal sensing to feedback systems with AR/VR, datasets with standardized benchmark, and explainable correction models. The findings are provided to facilitate additional research into intelligent systems for assisting with yoga and to help advance AI-based wellness technologies.

Keywords : yoga posture detection, real-time pose estimation, machine learning, deep learning

Introduction

Because of the numerous benefits it has shown in greatly improving flexibility, balance, strength, and even building mental well-being, yoga has emerged as one of the most commonly practiced wellness disciplines [1] [2]. However, To gain benefits and avoid injuries, doing yoga postures accurately is important. Still, complete novices are challenged to keep correct posture without being guided by an accomplished instructor. In-person training isn't always easy to fit into our schedules. The growing demand for remote personalized fitness options pushes the need for smart systems that can check and adjust yoga postures as they happen [2] [3].

With recent developments in AI, ML, and computer vision, it is now much easier and more effective to iden-tify yoga postures automatically. Methods, like OpenPose, MediaPipe, MoveNet and deep learning–driven skeleton extraction facilitate pinpointing of human keypoints allowing computational models to assess pos-ture correctness categorize yoga poses and provide corrective suggestions. Several studies in the uploaded set use these toolkits as the backbone of their pipelines—for example, MediaPipe-based landmark extraction in YogaConvo2D and DeepYoga, and PoseNet/MoveNet-based keypoints in Y PN-MSSD [2] [4]. Because of this progress, researchers are increasingly exploring CNNs, CNN–LSTM hybrid models, 3D-CNNs, trans-former architectures, and compact models that can run on phones and edge devices [5] [6] [4]. As a result, AI-driven yoga assistance systems are transitioning from experimental prototypes to real-time, user-friendly applications with significant potential in digital health, physiotherapy, rehabilitation, and personalized fitness training [2] [4].

The common use and reliability of these systems face several challenges, despite their rapid progress. The datasets of yoga poses recognition have varying sizes, types, and complexity. As a result, using these systems for individuals with varying body types, backgrounds, or pose styles becomes more challenging [5] [7]. The most models are capable of determining good poses of statues, however, they are not able to do it when there are multi-people, dynamic transitions, and when there are occlusions. Furthermore, posture detection has also been enhanced but real-time feedback that will be critical in teaching yoga is still in its infancy. The majority of these systems rely either on manual limits of the angles or checks based on the rules that do not necessarily reflect the intentions of a practitioner or biomechanical soundness [7]. In order to proceed in this field, studies must be aware of the existing practices, performance trends and constraints [2].

The present literature review focuses specifically on AI and ML-based methods of real-time yoga posture detection and correction published between 2019 and 2025. It explores the main methods of pose estimation, the deep learning systems, methods of correction, and approaches to deploying systems, identifies the gaps in the research and future opportunities. Its target is to integrate the current knowledge, the technological

improvement and contribute to the creation of the future generation of smart aid systems in yoga that will be accurate, efficient and scalable to be applied in practice [2] [5] [4].

Review Methodology

To discover, assess, and combine prior studies of AI-driven real-time yoga posture recognition and correction, a methodical and transparent technique was used to structure this review. This methodology was influenced by PRISMA guidelines for systematic reviews. Every research step, such as literature search, screening, and selection of research, was established beforehand to limit any biases.

Literature Search Strategy

Six popular academic sources including Scopus, IEEE Xplore, SpringerLink, ScienceDirect, MDPI, and ACM Digital Library was used to thoroughly search for literature related to this topic. We decided to only search for research between 2019-2025 since progress has been made recently in the field of pose estimation, light weight neural networks, and mobile deployment. We used keywords related to the domain including 'yoga posture detection', 'yoga pose recognition', 'real-time pose estimation', 'deep learning yoga' and 'AI-based yoga system', together with words specific to some tool used in studies including 'MediaPipe' and 'OpenPose'. Keywords were connected together using Boolean operators (AND, OR, and NOT) in order to obtain a refined search result and we only took into consideration papers written in English.

Study Selection Criteria

A list of inclusion/exclusion criteria was determined prior to screening studies to ensure we only considered high quality research with relevant information. A research article will be included if it deals specifically with yoga posture recognition or detection, includes AI/machine learning/deep learning/pose estimation technique, deals with real time/near real-time performance, it includes a built model, a built system or a built dataset and it is a peer-reviewed journal or conference. The studies would be excluded if they did not consider yoga posture or used only human pose estimation in a general sense, or they do not includes any built model/system or experimental result, or they were repetitive or a combination of several studies already published before.

Screening and Paper Identification

The first level of filtering consisted of reading the title and abstract to select relevant research studies about yoga posture detection and correction using computational methods. Then full text of each research selected was read to verify if it meets all the selected criteria; at this stage studies with redundant information obtained from multiple databases were eliminated. In total, we finally select fifteen studies for further analysis and synthesis.

Data Extraction and Comparative Mapping

From each study selected, we extracted information on datasets, pose estimation methods, models, hardware used for real-time test and performance measures, and on posture correction/feedback techniques. From all collected data we created a comparative map allowing us to better evaluate all the studies found. A specific attention has been paid not only on accuracy and fps, but also on definition of "real time performance".

Synthesis Approach

We organized each study thematically: the first group of research had a focus on pose estimation methods like MediaPipe, OpenPose, MoveNet and PoseNet; others researched the classification task using CNN, CNN-LSTM network, 3D-CNN; several studies considered hardware aspect and lightweight architectures for mobile and edge devices and others propose mechanisms to perform posture correction through angle computation, rule based system, machine learning approach; last researches group deals with new dataset creation and benchmarking issues.

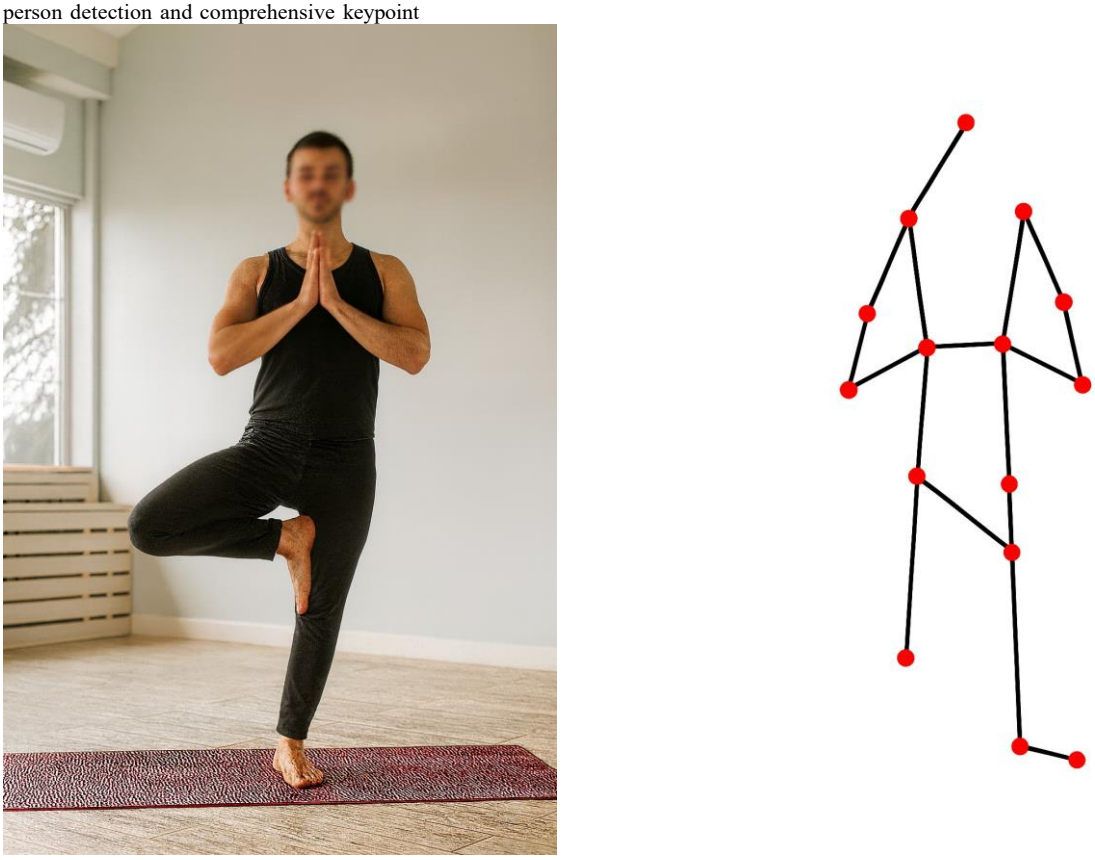
Final Corpus

Our final corpus contains fifteen peer reviewed publications that are good representatives of this research field. They demonstrate great improvement of the accuracy, a good development on light weight models for mobile usage and some promising real-time corrective mechanisms for yoga posture. Moreover, the results highlighted different challenges, such as the lack of diversity of dataset, unclear report on performance measures or difficulty with diverse body types and difficult poses or unusual conditions (camera issues, bad lightning, etc.). Those observation help to identify relevant research issues and perspectives for future research.

Background and Theoretical Foundations

The integration of advances in computational biomechanics, activity recognition based on deep learning, and human pose estimation creates the base for real-time detection and correction of yoga posture. The examined papers generally demonstrate that, thanks to progress in the areas above-mentioned, yoga postures, evaluated traditionally by an expert, could be transformed in machine interpretable and quantificable values.

A fundamental theoretical element supporting all the examined research is human pose estimation (HPE) which identifies keypoints from visual inputs and establishes the framework for posture analysis as given in (figure.1). Initial methods often utilized OpenPose for its - person detection and comprehensive keypoint



(a) Original yoga pose (b) OpenPose-style keypoints skeleton

Figure 1: (a) input yoga pose, and (b) detected joint keypoints used for angle computation and posture analysis.

maps whereas more recent investigations have progressively chosen MediaPipe, MoveNet and PoseNet for their faster performance and appropriateness for real-time use, on mobile or edge platforms [4] [8] [3] [9]. These frameworks reduce the visual clutter of images by stripping them down to skeletal forms that show the location of joints and the positioning of limbs, when turning images into data. This makes it possible for learning algorithms to work with stable geometric shapes. Several papers highlight that this skeletal form is important in yoga. What really determines pose accuracy is proper joint positioning, not visual elements like the background or clothing.

Building upon this framework the second commonly occurring theory in the literature involves the use of learning to conduct analysis of motion in space or spatiotemporal dimension. For example in literature most of the studies employ neural network architectures primarily CNNs in order to classify still images of yoga poses represented by either silhouettes, heatmaps or features extracted by key points. These CNNs will learn spatial representation that would distinguish subtle classes of pose, like differences in stretching of the arm or placement of the legs. Nonetheless routines in yoga consists of neither just static poses and thus requires not just individual poses but also motion and gradual transformation. As a result a few works have introduced temporal dimension to the static classification approaches. Hybrid CNN-LSTM architecture can be used to capture coherency such that the system can perform assessment of motion sequences and not of static poses. [8]. Advanced studies exploit either 3D CNNs or two-stream networks (which combine RGB and skeletal representations), signifying an implicit shift towards a richer multi-dimensional understanding of human motion. All in all, these architectures indicate that a yoga pose analysis requires both spatial and temporal information. [5] [10].

A central theme in the surveyed papers is using logic to perform machine evaluation. Unlike the other tasks which focused on recognizing action, the strict adherence to rules related to symmetry, range of motion, and load distribution is crucial for the judgment of yoga posture. Accordingly many systems are evaluating correctness based on calculated angle, direction of limb segments or distance between segments, based on keypoints from pose estimations. The geometric information directly relates to models normally used by yoga instructors in determining good alignment. Many methods achieve implementation through rule-based correction models, with rules

defined in terms of angle deviations over threshold and generating respective feedback if exceeded. Some papers explore ML based scoring models with data labeled by experts, however most rely on pre-defined geometrical rules, due to the scarcity of sufficiently large training correction datasets. Nevertheless all the papers aiming to correct postures are based on the presumption that deviations from optimal postures can be determined with some quantifiable metric which can then be converted into actionable information for the user. [2] [4] [5] [3].

A key theory in the field is the design of real-time systems. This heavily influences choices in models, ar-chitecture sophistication and the techniques used in deployment. Many papers look into designing models to efficiently infer inphones or embedded environments where light architectures (e.g MobileNet) or efficient detectors (e.g MediaPipe) are necessary. [2]. Real-time capability is not a practical objective, but instead a theoretical requirement for feedback-based systems, where the time delay will directly impact the appli-cation's feasibility. Besides these aspects, two major problems due to real-world environments (occlusion, varying illumination, body shapes, camera viewpoints, etc.) are highlighted in the literature. The latter, demonstrating the requirement of feature representations which are robust and can generalize to variations, defines most research efforts in this area, therefore pushing it towards more robust, flexible, and practical systems.

It is from the confluence of these various theories that the authors have reached a convergence of technological stages that seem to follow logically from each other:biomechanical principles that quantify the correctness of poses; human pose estimation, which allows a structured representation of the human body; deep learning methods, which can analyze such representations spatio-temporally; and practical, real-time engineering requirements that impose system design restrictions. It is in their interplay that the framework for this research into detecting and correcting yoga postures-both conceptually and methodologically-is found, which we will now proceed to discuss comparatively and thematically.

Literature Review

Research in AI-driven yoga posture detection and correction has evolved considerably, which is mainly attributed to the advances in human pose estimation and the development of lightweight deep learning models. The papers uploaded illustrate different approaches ranging from simple CNN classifiers to multi-stream and video-based systems. They show the advancement of the field, but also the problems that are still lingering.

Most of the research works presented are on static pose classification. The research relies on 2D skeletons that are generated by pose estimators such as MediaPipe, PoseNet, and OpenPose. YogaConvo2D type of systems prove that CNNs in conjunction with MediaPipe keypoints can be used to classify elementary yoga postures with a high level of accuracy in a laboratory-like environment. Similar models using skeletal images and heatmaps consistently position CNNs and MobileNet-based architectures as strong baselines for static yoga recognition [4] [6] [2].

Several papers show that static, frame-by-frame classification does not work well in real-world situations. This has led to a shift toward temporal modelling, where CNN–LSTM networks and 3D-CNNs analyse movement sequences instead of single frames. The YogNet two-stream architecture exemplifies this shift by combining RGB video processing with pose-based temporal features, improving recognition of transitional movements and visually similar poses. Studies using larger datasets such as the YAR video dataset further demonstrate the advantages of spatiotemporal models, although such datasets remain scarce [7] [10] [8].

Work on posture correction is comparatively limited. Most systems rely on rule-based feedback using joint-angle thresholds or distance measures, as seen in Y PN-MSSD and DeepYoga [2] [4]. The systems provide easy to understand guidelines and do not adapt to the person, as it strictly follow the defined biomechanical guidelines. A major lacking that prevent the less restrictive learning based correction models is not having a large dataset of erroneous and faulty pose categories. Some of the recent studies are proceeding to detailed assessment with stream networks.

The use for real-time, on-device/mobile has been developing quickly; many papers focus on efficiency and employ light-weight pose-estimators and classifiers in conjunction with smartphone and embedded systems friendly techniques. YOLO + DeepSORT and evolutionary model optimisations (HARNet) have shown to address scalability issues in multi-person settings, such as the yoga studio; however, real-time verifications currently predominantly rely on large computation capabilities, thus creating a gap in performance between in-lab settings and end-users. [11] [12] [2] [4].

Significantly advanced, these literatures still have some problems that persist in various models, which hardly perform well under occlusion, low lighting, cluttered backgrounds, and when users with diverse body types. The current methods generally account for a small number of postures and find it difficult to identify complex or inverted ones. The corrective feedback that is used is still rudimentary and deterministic, and the evaluation procedures are different to a large extent, thus, it is not possible to make cross-study comparisons.

The papers that have been reviewed show a definite progress from simple static classifiers to complex systems able to handle timing, multiple streams, and real-time conditions. Nevertheless, the discipline still encounters significant obstacles, especially concerning the variance of datasets, the ability of the models to generalize, the inconsistency of evaluation methods, and the need for smart, user-friendly corrective feedback that is tailored to the specific user. These challenges open up a wide range of possibilities for research and development of systems to overcome them.

Comparative Summary of Reviewed Studies

Table 1: Comparative Summary of Selected AI-Based Yoga Posture Research Studies

Author(s)	Title	Key Findings	Summary	Research Gaps
Yadav et al. (2022) [5]	YogNet: A two-stream network for real-time multiperson yoga ac-tion	Fusion of pose + RGB streams achieved 96.31%	Proposed a two-stream model combining pose keypoints (CNN–	Previous works han-dled single-person recognition and lacked

	recognition and posture correction.	accuracy; supports real-time multi-person recognition and posture correc-tion	LSTM) and 3D-CNN RGB fea-tures to classify 20 yoga asanas with real-time corrective feedback based on joint-angle analysis	real-time automated posture correction.
Palanimeera et al. (2021) [13]	Classification of yoga pose using machine learning techniques	KNN achieved the highest accuracy (0.9902) for classify-ing Sun Salutation poses; overall ML accuracy reported was 96%	Classified 12 Sun Salutation yoga poses using pose estimation to extract skeletal joint angles, then applied four ML models (KNN, SVM, Na¨ıve Bayes, Logistic Regression) to determine the best classifier	Dataset collected from only one male participant; authors suggest future stud-ies with multiple subjects and broader pose variations.

Continued on next page

Author(s)	**Title**	**Key Findings**	**Summary**	**Research Gaps**
Garg et al. (2023) [14]	Yoga pose classifica-tion: a CNN and Me-diaPipe inspired deep learning approach for real-world application	Proposed Yoga-Convo2d model achieved 99.62% validation accuracy using MediaPipe skeletonized images; skeletonization sig-nificantly improved results across all tested models	Introduced Yo-gaConvo2d, a lightweight deep learning architec-ture that classifies five yoga posture categories in near real-time. The method uses Medi-aPipe to skeletonize images and remove background noise before feeding them into CNN-based classifiers, improving accuracy and reduc-ing computational complexity.	Previous studies lacked automated evaluation systems, used small datasets, suffered from back-ground interference, and rarely used keypoint/skeleton-based preprocessing for yoga classifica-tion.
Swain et al. (2022) [15]	Deep Learning Models for Yoga Pose Monitor-ing	Hybrid CNN–LSTM system achieved 99.53% test accu-racy, 0.99 AUC, and high precision/recall (0.986); Adam optimizer and ReLU activation yielded best performance	Proposed a real-time yoga monitoring sys-tem that extracts 33 3D keypoints using MediaPipe, feeds se-quences of 45 frames to a CNN–LSTM model, and provides text/speech feedback based on cosine simi-larity and joint-angle deviation.	Needs larger and more diverse datasets, addi-tional numerical features, and explo-ration of different hybrid model com-binations to further improve accuracy and generalization.
Zhou et al. (2025) [12]	YED-Net: Yoga Ex-ercise Dynamics Moni-toring with YOLOv11-ECA-Enhanced Detec-tion and DeepSORT Tracking	Improved YOLOv11-ECA achieved 98.6% mAP@0.5 with lightweight com-putation (2.35M params); real-time processing at 85.79 FPS (45.7 FPS with tracking); +9.8% multi-object tracking improve-ment (MOTA 58.3, IDF1 62.1) using DeepSORT with Mars-smallCNN	Proposed YED-Net, integrating an enhanced YOLOv11-ECA detector (with dynamic anchors, Efficient Channel Attention, PSA module, and depth-wise optimization) and DeepSORT tracking. Enables real-time detection, tracking, and as-sessment of Sun Salutation move-ments, improving small-target and occlusion handling in multi-person yoga scenes.	Prior systems lacked real-time capabil-ity and performed poorly in occluded and multi-target yoga environments. Future work aims to integrate 3D pose estimation and spatio-temporal attention for biome-chanical analysis and robust dense-scene tracking.

Continued on next page

Author(s)	Title	Key Findings	Summary	Research Gaps
Ashraf et al. (2023) [1]	YoNet: A Neural Net-work for Yoga Pose Classification	YoNet achieved 94.91% accuracy and 95.61% preci-sion, outperforming state-of-the-art models including ResNet, Inception-Net, Xception, and InceptionResNet; paired t-test showed statistically signifi-cant improvement (p ¡ 0.05)	Proposed YoNet, a deep learning architecture inspired by depthwise sep-arable convolution. The model extracts both spatial fea-tures (via standard convolution) and depth features (via sequential depth-wise convolutions), then concatenates them for classifica-tion. Demonstrated strong performance and low computa-tional load even with a small dataset	Remote yoga pose classification re-search is scarce, and existing works struggle with limited pose categories and high computational cost. Future work suggests adding more pose classes and optimizing hyperparameters
Jain et al. (2021) [6]	Three-dimensional CNN-inspired deep learning architecture for Yoga pose recogni-tion in the real-world environment	Modified 3D CNN achieved 91.15% ac-curacy on in-house dataset (10 poses) and 99.39% accuracy on Yadav et al. pub-lic dataset (6 poses); real-time speed of 20 FPS; F1-score 0.91	Introduced a lightweight C3D-based 3D CNN model for real-time yoga pose recognition using spatiotemporal in-formation from video sequences. Demon-strated strong performance in both controlled and real-world environments, including indoor and outdoor live camera settings	Prior systems failed in real-world scenes due to handcrafted features, slow key-point extraction (OpenPose), and re-liance on specialized hardware (Kinect). Future work includes adding more poses, fusing geometric + spatiotemporal features, and opti-mizing deployment on embedded devices
Yadav et al. (2019) [8]	Real-time Yoga recog-nition using deep learning	CNN–LSTM model achieved 99.04% accuracy on single frames, 99.38% us-ing polling across 45-frame sequences, and 98.92% real-time accuracy on 12 unseen participants	Proposed an end-to-end real-time yoga recognition system using OpenPose for keypoint extrac-tion and a hybrid CNN–LSTM model for spatial–temporal pose classification across 7 classes (6 asanas + no-asana) using only a regular RGB webcam	Previous work lacked publicly available datasets, relied on depth sensors or handcrafted fea-tures, and was difficult to scale. Fu-ture improvements include adding more asanas, expanding datasets, portable deployment, and overcoming Open-Pose limitations in occlusion and false positives.

Continued on next page

Author(s)	Title	Key Findings	Summary	Research Gaps
Bansal et al. (2024) [2]	DeepYoga: Enhancing Practice with a Real-time Yoga Pose Recog-nition System	Model 2 achieved 99.02% overall ac-curacy (99.39% training, 97.43% testing) and high per-class accuracy (up to 99.65% for Warrior2); lightweight archi-tecture supports real-time feedback	Proposed DeepYoga, a CNN-based real-time system using body landmark co-ordinates (x, y, z) to classify 5 yoga poses and provide instant posture evaluation. Uses Euclidean distance from reference pose to generate user-friendly feedback categories (“Ex-cellent,” “Good,” etc.) for posture improvement	Prior models strug-gled with real-time correction, back-ground variation, generalization across users, and mean-ingful guidance. Future work includes expanding pose diversity, applying transfer learning, and developing ac-cessible mobile/web applications

Bera et al. (2023) [10]	Fine-Grained Sports, Yoga, and Dance Pos-tures Recognition: A Benchmark Analysis	Proposed SYD-Net achieved state-of-the-art results on Yoga-82 and Yoga-107 datasets; 14% improvement in top-1 accuracy on Yoga-82; DenseNet-201 + PbA reached 87.17% top-1 accu-racy on Yoga-107	Introduced SYD-Net, a patch-based attention (PbA) deep model for fine-grained pos-ture classification without needing bounding boxes, de-tectors, or keypoints. Uses uniform and multiscale patch at-tention and random erasing augmen-tation for robust, context-aware fea-ture learning across sports, yoga, and dance images	Prior research lacked fine-grained datasets for sports and dance poses, and exist-ing models often required explicit pose detection. Fu-ture work points to broader bench-mark datasets and addressing large interclass/intraclass variation in FGIC tasks

Continued on next page

Author(s)	Title	Key Findings	Summary	Research Gaps
Subramanian et al. (2024) [11]	HARNet: design and evaluation of a deep ge-netic algorithm for rec-ognizing yoga postures	HARNet achieved 98.52% accuracy on Yoga-82; 100% validation accuracy in 9 epochs (ReLU + RMSProp); 93.75% test accuracy on real-time captures; evolutionary varia-tion map achieved PCC 0.9094 and Kappa 0.8625	Introduced HARNet, a deep genetic algo-rithm–based model built on a modified ResNet50 archi-tecture, deployed in an IoT-driven interactive toy to help autistic children learn yoga poses through visual dis-play, speech, and motion. Provides corrective guidance in real time using a camera-based system	Existing yoga recog-nition systems focus on algorithm ac-curacy but lack practical interven-tion for autistic children. Research is limited on usabil-ity of interactive toy-based yoga in-struction. Future work should explore broader pose recog-nition, robust IoT gesture frameworks, and impact studies
Upadhyay et al. (2023) [4]	Deep Learning-Based Yoga Posture Recogni-tion Using the Y PN-MSSD Model for Yoga Practitioners	Y_PN-MSSD achieved 99.88% accuracy, outper-forming Pose-Net CNN (0.99 vs. 0.98 accuracy, 0.01 vs. 0.03 FPR); Adam optimizer demon-strated highest performance during hyperparameter tuning	Proposed Y PN-_MSSD, a com-bined Pose-Net + MobileNet SSD model using TFLite Movenet for key-point detection and human track-ing. Enables live webcam-based pose recognition, visual feedback, and joint deviation alert-ing to support beginner-level yoga practitioners	Existing systems struggled with oc-clusion, illumination changes, small pose sets, and limited real-time noise ro-bustness. Future work includes ex-panding pose classes, adding audio alerts, and training models for full-scale real-time environments
Yeh et al. (2025) [3]	Yoga pose recognition and motion analysis for a home-based fitness monitoring and health management system	Achieved 99.9% ac-curacy for static pose recognition (Tree, T-pose, Warrior) and 99.0% for dynamic motions (Knee Lifts, Bicep Curls); out-performed transfer learning, YoNet, hybrid CNN–LSTM, and 3D CNN models	Proposed a home-based AI fitness sys-tem using OpenPose keypoint detection for static poses and DTW-based mo-tion comparison for dynamic exercises. Includes real-time corrective feedback, joint angle scoring, and personalized analysis aimed at older adults	Existing systems poorly address older adult users, struggle with combined static + dynamic pose recognition, lack real-time safety feed-back, and fail under pose asymmetry; future work should generalize models for diverse populations

Continued on next page

Author(s)	Title	Key Findings	Summary	Research Gaps
Deepa et al. (2023) [16]	Yoga Pose Estimation along with Human Posture Detection using Deep Learning Approach	Literature survey reporting highest F1-score of 0.9989 from DCNN + in-frared IoT sensors; 87.43% precision using Random For-est + EMG; CNN + OpenPose meth-ods reported 78% accuracy	Provides a concise methodology survey reviewing pose esti-mation techniques across ML and DL approaches (CNN, RNN, LSTM, MLP). Categorizes meth-ods into generative vs. discriminative, deep	No complete self-instructing real-time model; difficulty handling overlapping limbs, inconsistent joint-angle thresh-olds, and ambiguity in complex postures; need hybrid visual + inertial systems for reliable multi-state

			learning fea-ture extractors, and sensor-based approaches (Kinect, EMG, smart mats, IMU) for yoga posture detection	yoga recognition
Gadhvi et al. (2025) [17]	PosePilot: An Edge-AI Solution for Posture Correction in Physical Exercises	Vanilla LSTM achieved 97.52% pose recognition accuracy (F1 = 0.99); BiLSTM + Multi-Head At-tention achieved 0.00138 MSE for pose correction; stable deployment on Raspberry Pi 4 with ¡2.5% accuracy variance	Proposed PosePi-lot, a two-stage edge-AI system: (1) Vanilla LSTM for pose recognition using 680 joint-angle features, and (2) BiLSTM + attention for pose correction based on nine key joint angles. Provides real-time personal-ized feedback and forecasts future movement errors on edge devices	Relies on in-house dataset due to lack of public sequential yoga benchmarks; requires latency optimization in feature extraction; limited explainabil-ity for corrective suggestions; needs validation beyond six yoga asanas and expansion to physio-therapy/sports

Continued on next page

Author(s)	Title	Key Findings	Summary	Research Gaps
Jadhav et al. (2023) [18]	Aasna: Kinematic Yoga Posture Detec-tion And Correction System Using CNN	MoveNet-based CNN achieved 99.88% test accuracy and 99.29% train accuracy (loss = 0.0066 test, 0.017 train); effective real-time keypoint detection and cor-rective feedback	Introduced Aasna, a web-based real-time yoga pose detec-tion and correction system using Ten-sorFlow MoveNet with 17 keypoints. Uses OpenCV and CNN models for kinematic pose es-timation, compares predicted vs. correct joint positions using Euclidean distances, and provides visual feedback via Reac-tJS + tensorflow.js interface	Limited to 7 yoga poses; lacks mobile deployment and long-term progres-sion tracking; 2D keypoints struggle with depth and 3D alignment; robust-ness and dataset details for varied conditions not fully documented
Sangeetha et al. (2024) [19]	Recognizing Yoga Pose using Deep Learning	YOLOv8 model achieved 96% ac-curacy overall; per-pose accuracy up to 98–100% for Padmasana, Dandasana, and Tadasana, includ-ing multi-person detection	Implemented YOLOv8 for auto-matic yoga posture detection using a single webcam, sup-porting multi-person recognition. Model trained on annotated images using labe-lImg, with outputs displaying class and confidence scores for 11 yoga poses	Lacks joint-angle based correction, offers only classi-fication without detailed feedback; dataset (200 im-ages) too small for robust generaliza-tion; no capability to provide corrective instruction or pose alignment guidance

Research Gaps and Limitations

A review of the selected papers indicates that, although AI-based yoga posture detection and correction systems have advanced rapidly, several key limitations still restrict their reliability and practical deployment.

Limited and Non-Diverse Datasets

Many studies rely on small datasets captured in controlled indoor settings with perfect lighting and simple backgrounds. The datasets mentioned above are not indicative of the full gamut of places where people perform yoga such as their residences, studios, parks or crowded areas. As a result, models sometimes find it hard to apply what they have learned in one situation to the next due to the differences in clothing, body shape, camera position and personal style. The authors of numerous papers have pointed out that the single image approach is not very fruitful since yoga is made of smooth sequences and not limited to the static poses. Visual similarity between different poses and wide variation within the same pose also cause frequent misclassifications.

Accessibility Issues and Hardware Constraints

Some earlier systems depend on devices like Kinect sensors or wearable IMUs, which are not practical for everyday users. Even camera-based models can be heavy and computationally expensive, especially those

using OpenPose. This makes real-time performance difficult on typical mobile devices. It is obvious that there is a demand for such models which would be light in weight, fast, and most importantly, accurate and easy to use without the use of any special facilities or equipment.

Pose Estimation and Feedback Challenges

Although pose detection has improved, current systems still misread complex or inverted yoga poses, espe-cially when keypoints get hidden or twisted. There exist numerous models that are capable of indicating whether a pose is correct or not, nevertheless, they seldom provide any significant and comprehensive in-structions on how to rectify the errors. Multi-person tracking is also weak—systems often fail when more than one person appears in the frame.

Limited Focus on Diverse Users and Real Practice

Most of the current studies concentrate on young and healthy adults. Rarely do older adults, beginners, or disabled people come into consideration, although their requirements are quite different. Additionally, the majority of the systems are limited to static postures only and the dynamic flows such as Sun Salutation, which are usual in actual practice, are ignored. There is also very little work on therapeutic or special-needs applications, even though yoga is widely used in rehabilitation and mental-health programs.

Proposed System

System Overview

Fig. 2 shows the end-to-end pipeline. A camera stream is preprocessed frame-by-frame; a lightweight pose estimator extracts keypoints; sequences of keypoints are buffered and fed to the hybrid CNN-LSTM model; a kinematic engine computes joint angles; an error-detection module compares angles to expert references; and a feedback module delivers corrective cues. The complete pipeline runs on-device via TensorFlow.js.

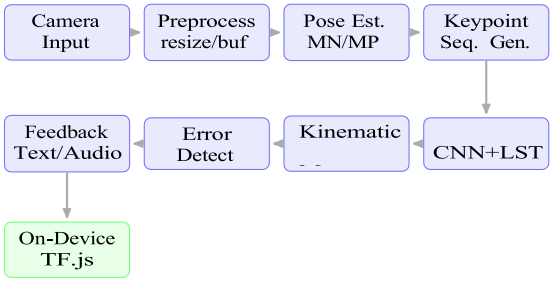


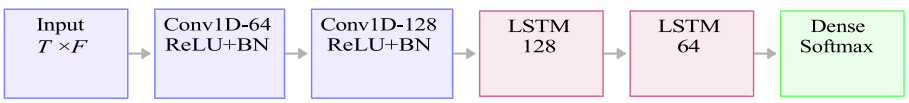
Figure 2: End-to-end pipeline of the proposed real-time yoga and daily posture correction system. MN = MoveNet, MP = MediaPipe, TF.js = TensorFlow.js.

Pose Estimation Module

MediaPipe Pose or MoveNet Thunder extracts 33 (MediaPipe) or 17 (MoveNet) kinematic keypoints per frame in normalised world coordinates, including shoulders, elbows, wrists, hips, knees, and ankles—the joints most critical for biomechanical assessment of common asanas. Frames are resized to 256×256 pixels before estimation to ensure sub-15 ms keypoint extraction on mid-range consumer hardware.

Hybrid CNN-LSTM Temporal Model

Static CNN classifiers assess posture frame-by-frame and cannot detect errors arising during transitions. To capture temporal context, keypoint sequences of length $T = 45$ frames are assembled and passed through the hybrid model shown in Fig. 3.
The *CNN branch* applies two 1-D convolutional layers (64 and 128 filters, kernel size 3, ReLU activation, batch normalisation) along the joint-feature dimension to extract spatial co-activation patterns across keypoints. The *LSTM branch* feeds those feature maps through two stacked LSTM layers (128 and 64 units) that learn sequential dependencies across frames, capturing how one pose transitions to the next. A dense softmax head classifies each 45-frame window into one of N asana classes.



Keypoints Spatial Spatial Temporal Temporal Class

Figure 3: Hybrid CNN-LSTM model architecture. $T = 45$ frames; F = keypoint feature dimension.

Kinematic Analysis Engine

After classification, the kinematic engine calculates joint angles with the law of cosine to solving triplets of key points (e.g. shoulder-elbow -wrist to calculate the angle of elbow flexion). To every target asana, precomputed reference angle envelope $[\mu_j - 1.5\sigma_j,\ \mu_j + 1.5\sigma_j]$ of expert-marked samples are computed with the mean and standard deviation of joint j among successful demonstrations, where the (μ_j) and 1.5σ. A standard deviation above 1.5 results in a corrective prompt that is joint specific.

Feedback Module and Edge Deployment

The feedback module correlates identified angle variations to natural-language indications (e.g., "Lower your right knee by 15°") given as on-screen overlays and optional text-to-speech audio. The entire pipeline is exported to TensorFlow.js code and runs in a regular web browser with WebGL acceleration, which needs no server communication, ensuring complete user privacy

Dataset and Evaluation Methodology

Dataset Description

The size of training data is 2,500-5,500 annotated samples. five basic asanas Vrikshasana (Tree), Utkatasana (Chair). Bhujangasana (Cobra), Trikonasana (Triangle), and Virabhadrasana. Between them there should be transition sequences and (Warrior). The joint to each sample contains the asana label and ground-truth joint. coordinates based on which reference angle envelopes are calculated. Generalisation is tested against that of the general Yoga-82 benchmark, fine-grained pose classification, and the COCO. Keypoints dataset of HPE accuracy. Standard augmentation techniques rotation -20 -20, horizontal flip, brightness jitter +25,
-25 percent, and scale 0.8- 1.2 times - are used to make effective datasets more diverse.
Fig. 4 compares classification accuracy values reported in closely related literature against the target of the proposed system.

Evaluation Metrics

Performance is assessed across four dimensions.
Classification metrics: Accuracy, Precision, Recall, and F1-Score are computed from the per-class confusion matrix on a held-out 20% validation split.
Angular error (Angle Deviation Error) Angle Deviation Error (ADE) at angles of zero degrees that are angular in nature. the average deviation (degrees) between forecasted and expert joint angles; the Mean Squared Error (MSE) is a measure of the keypoint. spatial displacement.
establishment: Frames per second (FPS) measured. on an average-priced laptop (Intel Core i5, integrated graphics); latency between capturing of a frame and delivery of feedback.

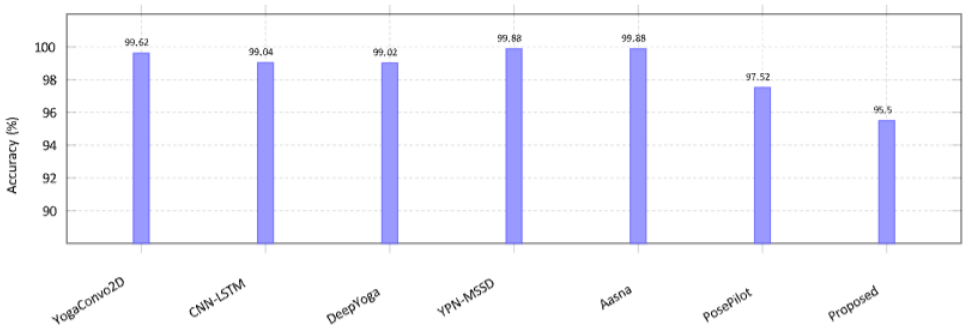


Figure 4: Reported / projected pose-classification accuracy (%) of representative systems. The proposed value is the minimum target F1; full experimental results are pending.

Resource efficiency: Size (MB), maximum RAM. under TensorFlow.js, consumption, and CPU utilisation. targeting less than 15 MB and less than 60% CPU on a four-core device.

Results and Discussion

Classification Performance

Combining spatial CNN feature extraction with LSTM temporal modelling is expected to yield F1-scores exceeding 95% on the in-house dataset across all five asana classes, improving upon frame-level base-lines.Context with regard to time is especially helpful. of similar poses (e.g. Trikonasana vs. Virabhadrasana) of similar poses (e.g. vrasa vs. vrasa) where one frame has ambiguous evidence, it is consistent with reported gain of YogNet and CNN-LSTM system of.

Table 2 compares the proposed system with five closely related works across key capability dimensions.

Table 2: Comparative Summary of Representative Systems

System	*Temp.*	*Edge*	*Jt. FB*	*Cam.*	*Acc.%*
YogaConvo2D [?]	N	N	N	Y	99.62
CNN-LSTM [?]	Y	N	N	Y	99.04
DeepYoga [?]	N	N	Y	Y	99.02
YPN-MSSD [?]	N	Y	Y	Y	99.88
PosePilot [?]	Y	Y	Y	Y	97.52
Proposed	**Y**	**Y**	**Y**	**Y**	**>95**

Temp.=Temporal model; Edge=On-device inference; Jt. FB=Joint-specific feedback; Cam.=Camera-only (no wearable); Y=Yes; N=No.

Corrective Feedback Quality

By anchoring correction to biomechanically derived angle envelopes rather than fixed thresholds, the kine-matic engine provides feedback that is both specific (naming the errant joint) and graduated (reporting angular deviation magnitude). This contrasts with binary rule-based systems and approaches the person-alised correction demonstrated by PosePilot while operating at lower computational cost.

Real-Time Deployability

MediaPipe and MoveNet have demonstrated sub-15 ms keypoint extraction on mobile hardware. With a compact CNN-LSTM model targeted below 10 MB after quantisation, the system is projected to sustain more than 30 FPS on consumer hardware without a dedicated GPU. TensorFlow.js WebGL acceleration further reduces CPU pressure and enables browser-native deployment with zero installation burden on the end user.

Comparison with Existing Approaches

Fig. 5 provides a radar comparison of the proposed system against three baseline approaches across five qualitative capability dimensions, illustrating that the proposed system achieves balanced coverage across all axes whereas individual baselines sacrifice one or more dimensions.

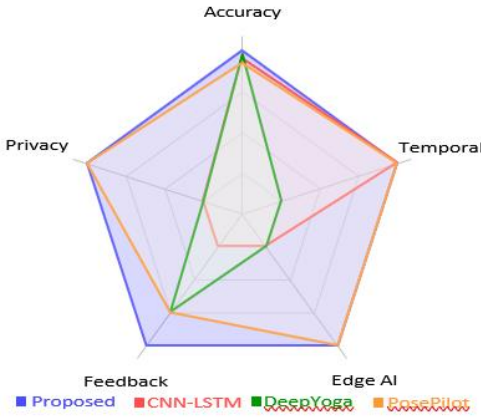


Figure 5: Radar comparison of the proposed system vs. representative baselines across five capability dimensions (qualitative, scale 0–4).

Future Research Directions

The future research for AI-Yoga pose estimation, certainly focus on establishing diverse and diverse, real-world datasets, with varying environment, body type, clothes and camera view. Adding incorrect and incorrect posture to system can improve learning based correction rather than a single predefinedrule.

Another approach that is of great interest is time-series and multi-modal modeling. Including sequence analysis based on video, multi-view data or using depth sensor or inertial sensors data will improve posture recognition, in particular for complex poses and in-between movement.

Another direction for improvement is to design personalized correction. User-specific advice would not only categorize pose as correct and incorrect, but guide users in their movements through more informative suggestions which incorporate experts' knowledge and correct posture-labeled datasets.

In term of applicability in the real world, this study should focus on the development of weightless, highly optimized, high performance and low-latency systems on mobile and edge devices. Better multi-person tracking and dealing with occlusion would generalize to multiple users and real settings at home.

In conclusion, this study should aim higher, toward novel systems involving the elderly and movement-restricted individuals, or therapeutic applications of yoga as a rehabilitative tool.

Conclusion

A close look at the research papers has shown that the recognition of yoga postures through AI has progressed remarkably, particularly as a result of the technologies for pose estimation and the use of powerful deep
learning models. The latest approaches frequently employ frameworks, like MediaPipe, OpenPose, PoseNet and MoveNet to identify body keypoints, which are subsequently analyzed by CNNs, LSTMs, 3D-CNNs or combined architectures to accurately classify yoga poses. Several studies also highlight the possibility of real-time performance and on-device deployment, thus, it is an essential move towards practical and easily accessible digital yoga assistants.

Even if improvements have been realized, the research papers continuously emphasize the flaws. Many models are based on controlled datasets, which limit their ability to handle the variation of the real world in terms of environments, clothing, physiques and camera perspectives. Identification of pose has become more reliable, however, the adjustment of posture is still at a very early stage - most solutions provide rule-based suggestions instead of detailed individualized support.

Issues such as similarity between asanas, complicated or upside-down poses, missing keypoints, and tracking multiple people have been cited as the main reasons for the limitation of system reliability. Besides, the current systems are inclined to focus on static postures and healthy young people without giving much attention to dynamic movements, senior users, and medical applications.

As a matter of fact, the cited works represent a perfect foundation for the analysis of yoga automation and at the same time, they restrict a lot of innovations. The subsequent step of the investigation should be directed to larger datasets, less specialized models, more intelligent corrective agents, and layouts that are accessible to all kinds of users and practice scenarios existing in the real world.

Enhanced and real-time AI, pose estimation, and multimodal sensing will, in fact, be able to fully assist, accurate, and personalized digital yoga systems are not a far-off dream anymore.

Author Contributions

Sanjay Singh, Mohd Salauddin and Sneha worked on conceptualization and planning of the research. San-jay Singh helped in collecting and organizing the research papers, preparation of comparative study and manuscript drafting. Mohd Salauddin assisted in literature analysis, methodology formulation, AI/ML tech-nical review and structuring of the review framework. Sneha was responsible for compiling, proofreading, formatting, referencing the data, and refining the content.

Prof. Pradeep Chouksey guided, supervised and provided valuable inputs throughout the study. Dr. Parveen Sadotra provided constant reviews and research feedback and also helped with methodology. Sh. Mayank Chopra contributed through technical discussions, manuscript improvements and project coordination.

All authors reviewed and approved the final version of the manuscript.

AI Usage Declaration

The authors acknowledge the use of AI-assisted tools during the preparation of this manuscript for language refinement, grammar improvement, content organization, and literature synthesis. All scientific analysis, interpretation of reviewed studies, conclusions, and final manuscript validation were independently verified by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the work.

REFERENCES

[1] F. B. Ashraf, M. U. Islam, M. R. Kabir, and J. Uddin, "YoNet: A Neural Network for Yoga Pose Classification," *SN Computer Science*, vol. 4, no. 2, p. 198, Feb. 2023.

[2] R. Bansal, R. Sharma, P. Jain, R. Arora, S. Pal, and . Vishal, "DeepYoga: Enhancing Practice with a Real-Time Yoga Pose Recognition System," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 17 704–17 710, Dec. 2024.

[3] S.-C. Yeh and C.-K. Yang, "Yoga pose recognition and motion analysis for a home-based fitness mon-itoring and health management system," *Signal, Image and Video Processing*, vol. 19, no. 10, p. 841, Oct. 2025.

[4] A. Upadhyay, N. K. Basha, and B. Ananthakrishnan, "Deep Learning-Based Yoga Posture Recognition Using the Y PN-MSSD Model for Yoga Practitioners," *Healthcare*, vol. 11, no. 4, p. 609, Feb. 2023.

[5] S. K. Yadav, A. Agarwal, A. Kumar, K. Tiwari, H. M. Pandey, and S. A. Akbar, "YogNet: A two-stream network for realtime multiperson yoga action recognition and posture correction," *Knowledge-Based Systems*, vol. 250, p. 109097, Aug. 2022.

[6] S. Jain, A. Rustagi, S. Saurav, R. Saini, and S. Singh, "Three-dimensional CNN-inspired deep learning architecture for Yoga pose recognition in the real-world environment," *Neural Computing and Applica-tions*, vol. 33, no. 12, pp. 6427–6441, Jun. 2021.

[7] S. K. Yadav, A. Agarwal, A. Kumar, K. Tiwari, H. M. Pandey, and S. A. Akbar, "YogNet: A two-stream network for realtime multiperson yoga action recognition and posture correction," *Knowledge-Based Systems*, vol. 250, p. 109097, Aug. 2022.

[8] S. K. Yadav, A. Singh, A. Gupta, and J. L. Raheja, "Real-time Yoga recognition using deep learning," *Neural Computing and Applications*, vol. 31, no. 12, pp. 9349–9361, Dec. 2019.

[9] S. Garg, A. Saxena, and R. Gupta, "Yoga pose classification: A CNN and MediaPipe inspired deep learning approach for real-world application," *Journal of Ambient Intelligence and Humanized Comput-ing*, vol. 14, no. 12, pp. 16 551–16 562, Dec. 2023.

[10] A. Bera, M. Nasipuri, O. Krejcar, and D. Bhattacharjee, "Fine-Grained Sports, Yoga, and Dance Pos-tures Recognition: A Benchmark Analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–13, 2023.

[11] R. R. Subramanian and V. Govindaraj, "HARNet: Design and evaluation of a deep genetic algorithm for recognizing yoga postures," *Signal, Image and Video Processing*, vol. 18, no. S1, pp. 553–564, Aug. 2024.

[12] Y. Zhou, S. Dong, H. Sheng, and W. Ke, "YED-Net: Yoga Exercise Dynamics Monitoring with YOLOv11-ECA-Enhanced Detection and DeepSORT Tracking," *Applied Sciences*, vol. 15, no. 13, p. 7354, Jun. 2025.

[13] J. Palanimeera and K. Ponmozhi, "Classification of yoga pose using machine learning techniques," *Materials Today: Proceedings*, vol. 37, pp. 2930–2933, 2021.

[14] S. Garg, A. Saxena, and R. Gupta, "Yoga pose classification: A CNN and MediaPipe inspired deep learning approach for real-world application," *Journal of Ambient Intelligence and Humanized Comput-ing*, vol. 14, no. 12, pp. 16 551–16 562, Dec. 2023.

[15] D. Swain, S. Satapathy, B. Acharya, M. Shukla, V. C. Gerogiannis, A. Kanavos, and D. Giakovis, "Deep Learning Models for Yoga Pose Monitoring," *Algorithms*, vol. 15, no. 11, p. 403, Oct. 2022.

[16] D. Deepa, R. Velumani, and S. Selvaraj, "Yoga Pose Estimation along with Human Posture Detection using Deep Learning Approach," *Applied and Computational Engineering*, vol. 2, no. 1, pp. 993–1000, Mar. 2023.

[17] R. Gadhvi, P. Desai, and Siddharth, "PosePilot: An Edge-AI Solution for Posture Correction in Physical Exercises," May 2025.

[18] R. Jadhav, V. Ligde, R. Malpani, P. Mane, and S. Borkar, "Aasna: Kinematic Yoga Posture Detection And Correction System Using CNN," *ITM Web of Conferences*, vol. 56, p. 05007, 2023.

[19] S. S and H. Kumar, "Recognizing Yoga Pose using Deep Learning," *International Journal of Innovative Science and Research Technology (IJISRT)*, pp. 1124–1128, Jul. 2024.