IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, November 2025
Impact Factor: 7.67

A Survey on AI-Based Real-Time Yoga Pose Detection and Correction Systems

Dr. Pradnya Bormane[1], Soham Ingale[1], Aditya Jagtap[1], Siddhant Kokate[1]
Department of Artificial Intelligence and Data Science[1]
AISSMS Institute of Information Technology, Pune, India
pradnya.bormane@aissmsioit.org, sohamingale2003@gmail.com,
adityajagtap7075@gmail.com, siddhantkokate26@gmail.com

Abstract: *The development of deep learning (DL) and artificial intelli- gence (AI) has revolutionized fitness technology by allowing real-time systems for correcting posture. Manual supervision is a major component of traditional exercise guidance tech- niques, which is frequently unavailable or inaccurate. Recent developments in computer vision, neural networks, and pose estimation frameworks have enabled intelligent systems with real-time feedback capabilities. This essay provides a thor- ough analysis of the main studies concentrating on the detec- tion and correction of yoga poses using AI. The study exam- ines the model architectures, datasets, accuracy, and constraints of current methodologies, including Vision CNN-SVM frame- works, MediaPipe, VGG16, Transformers (ViT), and Graph Neural Networks (GNNs). Lastly, it identifies research gaps and describes the path forward for developing a more reliable and user-friendly AI-powered yoga training system.*

Keywords*:* Artificial Intelligence, Posture Correction, Yoga Pose Estimation, Vision Transformer, MediaPipe, Deep Learn- ing, Computer Vision, GNN

I. INTRODUCTION

Due to its substantial advantages for both mental and physi- cal health, yoga has experienced a sharp increase in popular- ity worldwide [1]. But this increase, especially in at-home and self-directed practice, a major obstacle has been the possibil- ity of musculoskeletal harm from incorrect alignment and pos- ture. Practitioners may execute poses incorrectly without the in-the-moment supervision of a skilled instructor, which could cause strain and negate the practice's benefits [2, 1]. Accessible, precise, and automated guidance systems are now desperately needed as a result of this gap.

The domains of computer vision, deep learning, and artifi- cial intelligence (AI) have come together to tackle this problem. To develop "virtual yoga instructors" [3, 4]. Human pose esti- mation, a method for locating and monitoring the body's major joints (keypoints), forms the basis of these systems.From a reg- ular video feed. Technologies such as MoveNet, BlazePose, and Google's Me-diaPipe have become standard tools. For the real-time extraction of this skeletal data [2, 5, 6].

Keypoint extraction alone, however, is insufficient. Af- ter that, the system has to categorize the pose and, most cru- cially, determine whether it is correct. This survey looks at the wide variety of Deep Learning (DL) and Machine Learn- ing (ML) architectures created for this reason. We examine systems constructed on well-known Convolutional Neural Net- works (CNNs) such as VGG16 and EfficientNet [7, 8, 9], as well as hybrid models that combine CNNs with classifiers like Support Vector Machines (SVMs) [10]. Additionally, we look into new and emerging strategies, such as using Vision Trans- formers(ViT) [11], Graph Neural Networks (GNNs) [6], and the incorporation of Large Language Models (LLMs) to offer sophisticated feedback [12]. Additionally, this study examines strategies that address basic constraints like dataset scarcity by the development of extensive, multi-view datasets [13] and pri- vacy issues with camera-less WiFi sensing methods [14].

A thorough analysis of these cutting-edge techniques is given in this paper. By contrasting the approaches, performance indicators, and distinctive contributions of sixteen recent stud- ies, we provide a summary of the state of the field. The

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, November 2025
Impact Factor: 7.67

analysis of dynamic pose transitions and robustness to real-world envi- ronments are two examples of enduring research gaps and chal- lenges that we then highlight. We wrap up by suggesting future paths for the creation of AI-powered yoga training programs that are smarter, more reliable, and more customized.

The rest of this paper is structured as follows: The survey's methodology is described in detail in Section 2. A literature review organized by architectural approach is presented in Sec- tion 3. A discussion and comparative analysis of the surveyed papers are given in Section 4. Important research gaps are iden- tified in Section 5. Future research directions are outlined in Section 6, and the discussion comes to a close in Section 7.

II. SURVEY METHODOLOGY

The goal of this survey is to present a thorough analysis of the state-of-the-art in AI-based yoga pose detection and correction at this time. A systematic review procedure was used to accom- plish this.

Paper Selection

The majority of the papers in this review came from reputable online resources such as the CVF open access repository, IEEE Xplore, and other scholarly databases. To make sure the sur- vey includes the most recent developments, the search was re- stricted to recent publications from 2024 and 2025. ”Yoga pose detection,” ”yoga pose correction,” ”human pose estimation,” ”MediaPipe yoga,” ”MoveNet yoga,” ”Vision Transformer,” and ”Graph Neural Network” were among the keywords that were used.

Data Extraction and Synthesis

To enable a meaningful comparison, a standardized set of data points was extracted for each of the 16 chosen papers. Among the data were:

Core Model: Core Model: The main ML/DL architecture, such as CNN, ViT, GNN, and SVM.

Pose Estimation Technology: Pose Estimation Technology:

MediaPipe, MoveNet, and BlazePose are examples of key- point detection frameworks.

Key Contribution: Principal Contribution: The distinct nov- elty or emphasis of the paper (e.g., new dataset, feedback mechanism, or novel architecture).

Performance Metrics: Performance metrics include re- ported precision, accuracy, and other pertinent metrics.

Limitations: Identified shortcomings or difficulties, like light sensitivity, dataset size, or computational expense.

Following that, this data was combined to create the thorough comparative analysis that is shown in Section 4.

III. LITERATURE REVIEW

The field of study for AI-based yoga pose correction is broad, with a particular emphasis on enhancing precision, effective- ness, and practicality. Based on their fundamental architec- tural and methodological contributions, we classify the sur- veyed works.

CNNs and Hybrid Architectures

CNNs continue to be a mainstay of this field of study because of their demonstrated ability to extract features from images. In their comparative analysis of MobileNet, VGG19, and Ef- ficientNet with different optimizers, Tayal et al. discovered that EfficientNet with AdaDelta produced the best accuracy at 97.45 % [1]. Shirisha et al. achieved up to 98.82 % valida- tion accuracy after implementing a critical data cleaning step to remove corrupted images, demonstrating the popularity of the VGG16 architecture [2]. A real-time system aimed at pro- viding corrective feedback was also constructed by Syiemlieh et al. using VGG16 [3].

Hybrid models have also demonstrated potential. Dhole et al. achieved 98 % accuracy by combining VGG16 for fea- ture extraction with a Support Vector Classifier (SVC) [4]. In a similar vein, Mehta and Kaur created a CNN-SVM hybrid that achieved 89.17 % accuracy; however, they pointed out that there were issues with overfitting and confusion between similar pose classes [5].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, November 2025 Impact Factor: 7.67

Lightweight and Application-Focused Models

Efficiency has been given priority in a number of studies for real-world implementation on consumer devices. Using a Ran- dom Forest classifier, Dhanasekar and Sindhu developed a lightweight model that had a low computational overhead and a 90.17 % validation accuracy [6]. Optimized frameworks like MoveNet are frequently used in application-focused projects. Although its performance was sensitive to occlusions and inad- equate lighting, Rao et al. used MoveNet with MobileNetV2 to create a system with 98.3 % accuracy [7]. Beli et al. also achieved an impressive 99.97 % accuracy on correct poses by combining MoveNet with a neural network [8]. A MoveNet- based CNN was also utilized by Stephen et al., who achieved 98 % accuracy [9] by incorporating an angle calculation algo- rithm to provide voice feedback. Sikarwar et al. developed a custom CNN with MediaPipe, which achieved 87 % accuracy but highlighted the common issue of confusion between struc- turally similar poses [10].

Advanced and Novel Architectures

Some researchers have experimented with new architectures and data modalities, pushing the field's limits. The Vision Transformer (ViT) was first used by Wadawadagi et al. in their ”ViTYoga” system, which had a high Average Precision of 92.88 % but had trouble with fine-grained pose variations [11]. ”YogaGNN,” a Graph Neural Network that represents the hu- man skeleton as a topological graph, was presented by Zhang and YiMin [12]. It outperformed standard CNNs in terms of training efficiency and achieved 86.19 % accuracy on the large Yoga-82 benchmark [12].

Other innovations concentrate on feedback and data. Chauhan et al. created a more reliable correction system by analyzing temporal data from 3D landmarks using a hybrid CNN-LSTM model [13]. To provide more individualized and nuanced feedback, Chittimalla and Potluri combined a ResNet model with Large Language Models (LLMs) [14]. M3GYM, a large-scale, multi-view, multi-person dataset for fitness activ- ities, was introduced by Xu et al. to address the fundamental problem of dataset limitations. This dataset will be essential for training models that can handle real-world occlusions and group settings [15]. Lastly, in order to address privacy concerns, Gorrepati et al. proposed ”YogaFi,” a camera-less system that recognizes poses using WiFi Channel State Information (CSI) signals with an accuracy of 91 % [16].

IV. COMPARATIVE ANALYSIS

Table 1 offers a thorough synopsis of the 16 research papers in order to synthesize the reviewed literature. The table illus- trates the range of strategies, including WiFi-based techniques, experimental GNNs, and well-known CNNs. The main con- tribution or limitation of each study is contrasted with the re- ported performance metrics, the fundamental model architec- tures, and the underlying keypoint detection technologies (e.g., MediaPipe, MoveNet).

Several significant trends emerge from the data analy- sis. First, it is consistently possible to achieve high accuracy (above 90 %). Several models have achieved 98 % or higher [4, 2, 7, 9, 8]. Second, there is frequently a trade-off when se- lecting a keypoint detector: MediaPipe and MoveNet are pre- ferred due to their real-time performance on portable devices [6, 7], whereas component-based systems may employ more sophisticated models. Third, a clear distinction is emerging between models focused purely on classification accuracy and application-focused systems that prioritize real-time feedback [9, 6, 13] and privacy [16].

V. RESEARCH GAPS AND CHALLENGES

Even with the encouraging outcomes shown in Table 1, a num- ber of significant and persistent obstacles stand in the way of creating a truly reliable AI yoga instructor.

Limitations in datasets are a major problem. The ability of many models to generalize to new users, body types, and en- vironments is compromised because they are trained on small, clean, or simple datasets (like Kaggle datasets) with a limited number of poses [10, 6]. Because it captures the complexity of real-world settings, the development of new, large-scale, multi- view, and multi-person datasets like M3GYM is an important step forward [15].

Additionally, the systems are extremely susceptible to envi- ronmental influences. Body occlusions (both self-occlusion and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, November 2025
Impact Factor: 7.67

Table 1: *Comparative Summary of Reviewed Research Papers on Yoga Pose Detection.*

Author(s), Year	Model Architecture	Keypoint Tech	Accuracy (%)	Key Contribution / Limitation
Wadawadagi et al. [11]	Vision Transformer (ViT)	COCO-Pose	92.88 (AP)	Novel use of ViT; struggles with fine-grained pose variations.
Dhole et al. [4]	VGG16 + SVC	MediaPipe	98	Effective CNN-SVM hybrid approach for high accuracy.
Dhanasekar & Sindhu [6]	Random Forest	MediaPipe	90.17 (Val)	Lightweight model with audio feedback; weak on dynamic poses.
Mehta & Kaur [5]	CNN-SVM Hybrid	Custom CNN	89.17	Notes issues with overfitting and class confusion.
Shirisha et al. [2]	VGG16	Custom CNN	98.82 (Val)	Emphasizes rigorous data cleaning for high performance.
Zhang & YiMin [12]	Graph Neural Network	BlazePose (3D)	86.19	Innovative GNN approach; more efficient to train than CNNs.
Rao et al. [7]	MobileNetV2	MoveNet	98.3	High accuracy; sensitive to poor lighting and occlusions.
Stephen et al. [9]	CNN	MoveNet	98	Integrated angle calculation algorithm with voice feedback.
Sikarwar et al. [10]	CNN	MediaPipe	87 (Val)	Highlights confusion between structurally similar poses.
Tayal et al. [1]	EfficientNet	Custom CNN	97.45	Comprehensive comparison of CNNs and optimizers.
Syiemlieh et al. [3]	CNN (VGG16)	MediaPipe	N/A	Focuses on real-time system with corrective feedback.
Chauhan et al. [13]	Hybrid CNN-LSTM	MediaPipe (3D)	91.7	Uses 3D landmarks and temporal data for correction feedback.
Beli et al. [8]	MoveNet (NN)	OpenCV	99.97	Real-time web app; 99.97% on correct poses, 83.92% on errors.
Chittimalla & Potluri [14]	CNN (ResNet) + LLM	Yoga-82	92	Novel use of Large Language Models (LLMs) for nuanced feedback.
Gorrepati et al. [16]	Random Forest / XGB	WiFi Sensing (CSI)	91	Privacy-preserving approach using WiFi signals, not cameras.
Xu et al. [15]	N/A (Dataset Paper)	N/A	N/A	Introduces M3GYM: a large-scale, multi-view, multi-person dataset.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, November 2025 Impact Factor: 7.67

occlusion by others), background clutter, and inadequate light- ing can all considerably impair performance [7, 15]. This limits the real-world usability of many proposed systems.

Additionally, the majority of existing models are very good at categorizing static poses, but they are unable to examine the dynamic transitions between them, which is a crucial compo- nent of yoga flow (Vinyasa) [6]. This is partially addressed by models such as Chauhan et al. [13]'s CNN-LSTM hybrid, but there is still a sizable gap.

Lastly, there are still two real-world obstacles: privacy and computational cost. Real-time deployment on accessible devices, such as smartphones, is hindered by the computational cost of complex deep learning models [12].

VI. FUTURE DIRECTIONS

Future research should focus on a few crucial areas in order to progress the field and get past the obstacles mentioned above. The first is the creation and uptake of extensive video datasets such as M3GYM, which document a diverse range of postures, practitioners, and settings, including complete yoga sequences in multi-person settings [15].

Additionally, feedback quality needs to be improved. Say- ing ”correct” or ”incorrect” is not very useful. A crucial next step is integrating Large Language Models (LLMs) to convert intricate geometric data into insightful, tailored, and motivating feedback [14]. When paired with Explainable AI (XAI) con- cepts, this can assist users in comprehending *why* a pose is wrong and *how* to correct it.

Last but not least, research must keep looking into model optimization for edge devices and privacy-preserving strategies for deployment that is both realistic and moral. One crucial area of research for developing systems that users can rely on in their private spaces is the use of non-visual sensors, such as the WiFi (CSI) method [16].

VII. CONCLUSION

Sixteen recent research papers on the creation of AI-driven sys- tems for yoga pose detection and correction have been compiled and examined in this survey. According to our research, the field is developing quickly, progressing from basic classification to complex, real-time feedback systems. Convolutional neural networks, especially those with architectures like VGG16 and EfficientNet, have been clearly seen as the mainstay for fea- ture extraction and classification. These are commonly used in conjunction with portable, fast keypoint detectors such as Medi- aPipe and MoveNet, enabling high-accuracy (often ¿98%) per- formance in controlled settings.

But this survey also highlights a number of important, enduring issues that need to be resolved. Many models' high performance is frequently based on small, clean datasets. When confronted with real-world complexities like dim lighting, background clutter, and bodily occlusion—problems that are only now being tackled by new, extensive datasets—this re- sults in a robustness gap. Furthermore, the crucial spatiotem- poral dynamics of pose transitions—a crucial aspect of yoga practice—are largely ignored by the great majority of current research, which concentrates on static pose classification.

In the future, this field's trajectory suggests that these par- ticular issues will be resolved. We predict that static models will give way to more intricate architectures such as CNN-LSTM hybrids and Graph Neural Networks, which are able to interpret not only the ”what” of a pose but also the ”how” of a movement. At the nexus of modalities, the biggest innovations will probably take place. A shift toward more intelligent, moral, and user-centered systems is indicated by the incorporation of Large Language Models for producing genuinely personalized and instructive feedback as well as the investigation of privacy- preserving modalities like WiFi sensing. Ultimately, the goal is to transcend simple pose ”detection” and create a virtual assis- tant that captures the nuance and safety of a human instructor, making the benefits of yoga practice truly accessible to all.

REFERENCES

[1] D. K. Tayal, H. Dattatrey, A. Nagpal, R. Arya, A. Jain, and H. Gaur, “Comprehensive Analysis of Deep Learning Approaches for Yoga Pose Detection,” in 2025 International Conference on Innovation in Computing and Engineering (ICE), 2025.

[2] G. Shirisha, N. R. Bhat, A. S. Hamasagar, A. A. Hosamani, and P. Patil, "An Improved Approach for Yoga Pose Estimation of Im- ages," in 2024 5th International Conference for Emerging Tech- nology (INCET), Karnataka, India, 2024, pp. 1–7.

[3] B. Syiemlieh, L. Jyrwa, C. C. Marak, and S. K. Chettri, "Real- Time Yoga Pose Detection and Classification Using CNNs for Enhanced Fitness and Rehabilitation," in 2025 4th OPJU Inter- national Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0, 2025.

[4] S. A. Dhole, A. Galande, S. Chaudhari, and P. Dhembare, "Yoga Pose Detection using Deep Learning," in IEEE Proceedings, 2023.

[5] S. Mehta and S. Kaur, "From Mats to Models: Applying CNN- SVM Hybrid Techniques for Yoga Pose Classification," in 2024 3rd International Conference for Advancement in Technology (ICONAT), Goa, India, 2024, pp. 1–6.

[6] S. Dhanasekar and S. Sindhu, "Asana AI - Yoga Pose Detection Using Pose Estimation and Real-Time Feedback System," in 2025 International Conference on Computational Robotics, Testing and Engineering Evaluation (ICCRTEE), 2025, pp. 1–6.

[7] S. Rao, Nishmitha, A. D. Bhat, C. T, D. D. N, and Rakshitha, "TechAsanaAdvancer: Technical Mastery of Yoga Poses with AI," in 2024 Second International Conference on Advances in In- formation Technology (ICAIT-2024), 2024, pp. 1–7.

[8] S. Beli, B. Patel, and Mohana, "Real Time Yoga Pose Recogni- tion and Classification using MoveNet Deep Learning Model and Computer Vision," in 2025 6th International Conference on In- ventive Research in Computing Applications (ICIRCA), 2025, pp. 1829–1834.

[9] S. Stephen, J. James, L. M. Paul, P. Priyesh, and S. Vinod, "Fit- ness Help: Yoga Pose Estimation and Correction with Voice Feed- back," in 2025 Advanced Computing and Communication Tech- nologies for High Performance Applications (ACCTHPA), 2025, pp. 1–6.

[10] K. Sikarwar, A. Dubey, N. Venu, and D. Bisen, "AI-Based Sys- tem for Real-Time Yoga Pose Detection and Correction," in 2025 IEEE International Students' Conference on Electrical, Electron- ics, and Computer Science (SCEECS), 2025.

[11] R. Wadawadagi, V. E, A. V. G, V. Desai, and S. B. Patil, "Vi- TYoga: Vision Transformer for Real-time Yoga Pose Estimation and Analysis," in 2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF), 2025, pp. 1–6.

[12] M. Zhang and Q. YiMin, "Yoga Pose Recognition Method based on Invariant Graph Neural Network," in 2025 37th Chinese Con- trol and Decision Conference (CCDC), 2025, pp. 3397–3402.

[13] V. Chauhan, M. Aggarwal, Y. Singhal, and S. Shukla, "Real-Time Yoga Pose Recognition and Correction Feedback System Using Deep Learning and Computer Vision," in 2025 2nd International Conference on Computational Intelligence, Communication Tech- nology and Networking (CICTN), 2025, pp. 844–849.

[14] S. K. Chittimalla and L. K. M. Potluri, "Enhancing Yoga Pose Recognition using Large Language Models with Real-World Dataset Integration for Personalized Fitness Assistance," in 2024 International Conference on IT Innovation and Knowledge Dis- covery (ITIKD), 2024.

[15] Q. Xu, R. Cao, X. Shen, H. Du, S. Wang, and X. Yu, "M3GYM: A Large-Scale Multimodal Multi-view Multi-person Pose Dataset for Fitness Activity Understanding in Real-world Settings," in 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025, pp. 12 289–12 300.

[16] G. Gorrepati, S. P, and S. K. Udgata, "YogaFi: WiFi sensing based Yoga Pose Recognition using Machine Learning Model," in 2024 17th International Conference on Sensing Technology (ICST), 2024.