Sport Sciences for Health
https://doi.org/10.1007/s11332-025-01328-7

RESEARCH

Angle-based regularized deep learning model for gauging effectiveness in performing yoga postures

Akshansh Rawat[1] · Ananthakrishnan Balasundaram[2] · Chockalingam Aravind Vaithilingam[3]

Received: 19 June 2024 / Accepted: 7 January 2025

Abstract
Practicing yoga benefits both mental and nervous system and helps mitigate several health problems. However, not performing yoga in the correct manner can worsen the symptoms as well. This work presents a novel technology-driven approach named spondylitis-related yoga MediaPipe angle-based regularized network (SpY_MARNet) to assist people to perform yoga postures for treating spondylitis in the correct way. This work enables real-time interaction and provides immediate feedback, assisting in correcting postures and suggesting modifications if necessary. This work also monitors the duration for which each pose is retained which is a vital aspect of yoga practice. Also, the users are categorized into beginner, intermediate, and advanced levels based on their yoga performance. By using the model we achieved an accuracy of 99.7%. The results indicate significant promise in aiding individuals with spondylitis, opening avenues for further research and application in other physical therapies and wellness practices.

Keywords SpY_MARNet · Spondylitis · Random forest · Deep learning · Pose estimation · Pose correction · Pose recognition

Introduction

With the growing popularity in practice of yoga, recent epidemiological research has highlighted the injuries that are associated with yoga by utilizing data from the National Electronic Injury Surveillance System (NEISS) [1]. It uncovers an upward trend in yoga injuries with participants of age more than 65 years. People doing yoga independently further increase the risk of injuries. This background sets the stage for our research, which aim to enhance the safety and effectiveness of yoga practice.

Some yoga posture soothes the symptoms of spondylitis, helping participants in flexibility, endurance, strength, and mental peace. However, the effectiveness of yoga largely depends on how well the participants is performing yoga and the time duration for which the position has being held. There is a necessity to not only let the participants know about the correctness of yoga posture but also the time-duration for which he has held the pose. The integration of artificial intelligence (AI) in monitoring and guiding yoga practice presents a transformative solution. AI-driven systems that can provide real-time feedback for every wrong posture to ensure yoga is performed with maximum accuracy. There is also a need for enhanced safety and effectiveness in yoga practice to be addressed by introducing a cutting-edge computer-vision-based methodology. When performing any yoga postures, it is important that the users are free from wearing any additional wearables or hardware as it hampers their mobility. This makes vision bases systems a good choice as the user need not wear any additional gadgets while doing yoga. While IMU sensors do offer advantages like high precision in motion tracking and robustness to environmental lighting conditions, they come with limitations such as the need for multiple devices to be worn, potential discomfort, and additional costs for hardware. The vision-based approaches, on the other hand, leverage widely available devices like cameras, which significantly lower the barrier to entry. A comparative analysis between these modalities would highlight the trade-offs: vision-based systems excel

✉ Ananthakrishnan Balasundaram
balasundaram.a@vit.ac.in

1 School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India

2 Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai 600127, India

3 Clean Technology Impact Lab, Taylor's University, 47500 Selangor, Malaysia

Published online: 27 February 2025

in accessibility and scalability, while IMU sensors provide direct, reliable motion data but at the cost of usability and affordability. This trade-off aligns with our goal of creating a practical and inclusive solution for assessing yoga postures. Future work could explore hybrid approaches to combine the strengths of both modalities for enhanced accuracy and user experience.

A deep learning novel convolutional neural network (CNN)-based approach for human pose estimation which is distinctive in its creation of part heatmaps for individual body parts, using a cascading CNN structure [2]. The approach enhances the precision of pose estimation, crucially focusing on the subtleties of each body part where initially body parts are detected via a part detection network, and subsequently refining these detections through a regression sub network. Crucially, this method addresses and overcomes the limitations inherent in holistic models, offering more accurate and granular pose estimations, particularly in images characterized by occlusions and complex interactions. This advancement in pose estimation technology is particularly pertinent to our work, as it aligns with our goal of developing a system that provides nuanced and accurate feedback on yoga poses, catering to the diverse needs and conditions of individuals. But the CNN-based approach, especially one that involves cascading networks for partial heatmap generation and refinement, can be computationally more intensive. This might require more powerful hardware and could be less efficient in terms of processing speed, making it less suitable for real-time applications or devices with limited computational resources.

The proposed system involves a real-time yoga posture detection model specifically designed for spondylitis treatment with utilization of the Mediapipe framework [17] for posture detection and analysis by converting skeletal data into angles and distances for posture evaluation. Using the posture evaluation to a machine learning algorithm that evaluates the correctness of poses, tailored for different skill levels and along with it recording and monitoring the duration of each yoga pose and providing immediate feedback on the accuracy and duration of poses thereby enhancing the effectiveness of yoga practice for individuals with spondylitis. This work leverages computer vision to provide real-time feedback for yoga poses. This technology aims to create a safer and more effective yoga experience. To address privacy concerns, the system extracts only essential features like joint landmarks without storing or transmitting raw images. This approach minimizes the risk of sensitive visual data being exposed. To mitigate low quality visual sensing, Mediapipe, a robust framework for pose estimation, known for its ability to work effectively under varied conditions, including low-resolution cameras and suboptimal lighting is used. This choice helped mitigate the impact of poor visual quality to a significant extent. To ensure compatibility across a wide range of devices, the system was optimized for real-time performance and low computational overhead. By relying on lightweight models and efficient preprocessing, the framework remains accessible even on devices with limited hardware capabilities. By tailoring the practice to individual needs, it can minimize injury risk and maximize the therapeutic benefits of yoga. This research endeavors to merge the ancient wisdom of yoga with contemporary AI technology, offering an innovative solution for managing spondylitis. The implications of this study extend beyond spondylitis management, potentially revolutionizing the way we approach physical therapy and wellness through technology-enhanced, personalized care. By harnessing AI's capabilities in accurately assessing and guiding yoga postures, this research not only provides a supportive tool for individuals with spondylitis but also paves the way for broader applications in various musculoskeletal conditions and rehabilitation scenarios. This fusion of traditional practices with modern technology holds promise for improving quality of life and health outcomes, democratizing access to specialized therapeutic guidance, and fostering a deeper, more informed engagement with personal health and wellness routines.

The main contributions in this work are listed below:

- This work integrates AI with traditional yoga practice by fusing advanced AI technologies, such as Mediapipe for body landmark extraction, with the traditional practice of yoga to provide a novel approach for health and wellness, particularly benefiting individuals with spondylitis has been proposed.
- This work promotes real-time pose correction by providing immediate feedback on yoga poses performed thereby helping the users to perform the yoga posture in the right way, which is critical in preventing any ill effects and ensuring the effectiveness of yoga practice.
- This work monitors the duration and accuracy of yoga poses over time and accordingly the users are categorized into different proficiency levels (beginner, intermediate, and advanced).
- The proposed work measures the accuracy and the endurance of users while performing yoga poses using a new metric. It also quantitatively measures the improvement over time in performing this yoga poses. By comparing current performance metrics—such as pose accuracy, hold duration, and overall session quality—with historical data, the system can provide valuable insights how much user has improved over time. This encourages users by highlighting their progress, areas of improvement, and setting achievable goals for further advancement.
- To further encourage user engagement, the system visualizes improvement using line and bar graphs, offering valuable feedback and boosting motivation.

The manuscript is structured such that section "Related work" walks through the various contemporary approaches carried out towards yoga posture recognition. Section "Proposed system" details the proposed work while section "Experimental results and analysis" discusses in detail about the experimental results. Finally, section "Conclusion and future work" forms the conclusion section that discusses the highlights of this work and the scope for future expansion of this work.

Related work

This section discusses about a few contemporary works related to yoga posture recognition. The work in [3] exemplifies the integration of deep learning in the health and wellness sector. The development of a self-assistance system for real-time identification and correction of yoga postures and hand mudras leverages the YOGI dataset. By utilizing skeleton models for feature extraction and machine learning models, notably XGBoost with RandomSearch CV, for recognition, the system achieves an impressive accuracy of 99.2%. This innovation is further extended to a mobile application, transforming it into a virtual yoga instructor, thus enhancing yoga practice through technological means. In a parallel development, a novel approach highlighted in [4] introduces a deep learning model adept in real-time camera pose estimation. Employing convolutional neural networks to analyze single RGB images, PoseNet excels in diverse settings, from indoor environments to large-scale outdoor scenes. Its robustness against various challenges, including variable lighting and motion blur, underscores its potential applications in augmented reality and robotics, with notable accuracy metrics in different scenarios. The work [5] further expands the scope of real-time recognition technology. By adopting the motion history image (MHI) technique, this system effectively captures and analyzes the temporal aspects of hand gestures. Demonstrating high efficiency and accuracy, with an average accuracy of 94.1% and a processing time of 3.81 ms per frame, this model sets a benchmark in gesture recognition, applicable in real-time environments.

The work [6] introduces a system for recognizing sign language words using the Kinect sensor and a multi-stream hidden Markov model (MS-HMM). The approach focuses on analyzing hand movements, positions, and shapes to interpret sign language. It uses two datasets for Japanese and Italian sign languages. The methodology is innovative in integrating Kinect's capabilities with MS-HMM, demonstrating the potential for enhanced sign language recognition. The paper details the effectiveness of this approach in bridging communication gaps for the deaf and hard of hearing communities. The system achieves an overall efficient performance in sign language recognition tasks, demonstrating the efficacy of the combined approach.

The authors of work [7] employs the tf-pose-estimation algorithm, creating a custom dataset, YOGI, with images of ten yoga poses. The methodology includes extracting skeletal joint angles and applying machine learning classifiers for pose identification. The study focuses on enhancing the accuracy of pose recognition, contributing to the field of smart health care and personal fitness. The research presents a novel approach to integrating technology with traditional practices like yoga. The best-performing model, Random Forest classifier, achieved an accuracy of 99.04%.

A comprehensive overview of methods for evaluating human actions using vision-based technologies is provided by [8]. The paper discusses the evolution of these methods, categorizing them into traditional handcrafted feature-based methods and contemporary deep learning approaches. It highlights key challenges in action evaluation such as dealing with complex backgrounds and varying lighting conditions. The survey also covers various benchmark datasets used in this domain and performance evaluation metrics. The paper is valuable for understanding the breadth of methodologies and their applications in different contexts, offering insights into the strengths and limitations of current approaches in vision-based human action evaluation.

Authors of [9] introduced an advanced method for analyzing and evaluating fitness exercises using 2D and 3D skeleton data. This approach employs a unique dataset, Fitness-28, which includes various fitness actions. The methodology involves precise skeleton data extraction and processing, feature encoding, and the use of sophisticated classifiers like SVM and CNN for action analysis. This comprehensive framework allows for the effective evaluation of fitness actions, offering insights into the quality and correctness of exercises performed. The approach achieves high accuracy, with an average of 97.24% in front view recognition on the Fitness-28 dataset for SVM.

A novel approach to gait [10] recognition, focusing on handling both known and unknown covariate conditions. The methodology involves a combination of convolutional neural networks (CNN) and a discriminative feature-based classification method. The research emphasizes addressing challenges in gait recognition due to variations like clothing and walking speed. The study employs the CASIA and OUR-ISIR gait datasets for analysis and validation. This approach demonstrates significant advancements in gait recognition under varying conditions, achieving high accuracy rates, and offering insights into handling unknown covariates in gait analysis. The reported accuracy is 90.32% for the CASIA dataset under unknown covariates.

A novel method for real-time human pose tracking tailored for mobile devices is provided in [11]. The paper introduces a lightweight CNN architecture, BlazePose, capable of running efficiently on-device. It uses both heatmap-based and regression-based approaches to detect 33 body keypoints. The model was trained and evaluated using a dataset specifically curated for common poses and fitness exercises. BlazePose offers a significant advancement in achieving a balance between high accuracy and real-time performance on mobile devices, addressing the challenges of computational limitations and the need for precise pose tracking in various applications. BlazePose achieves a PCK@0.2 score of 97.2% after re-annotation of AR dataset independently.

The work in [12] presents a methodology for yoga posture recognition using Microsoft Kinect. It focuses on real-time detection of human joint points to evaluate the accuracy of yoga poses. The proposed approach utilizes Kinect to capture the 3D coordinates of body joints, calculating angles between them to assess pose correctness. This system offers a novel means for monitoring and guiding yoga practice, especially useful in fitness and rehabilitation contexts. The paper demonstrates the potential of Kinect in enhancing the precision and effectiveness of yoga training.

The work in [13] presents a novel approach to human pose estimation, focusing on structured feature learning. Their methodology involves reasoning correlations among body joints at the feature level, a departure from traditional methods that focus on score maps or predicted labels. The researchers introduce geometrical transform kernels implemented within a convolution layer, enabling joint learning of features, and their relationships. They propose a bi-directional tree structured model for optimizing feature channels at each body joint, significantly improving feature learning. The framework demonstrates substantial improvement in mean PCP on the LSP and FLIC datasets, showcasing its effectiveness in pose estimation.

To deal with such computational issue, there is a need to build further upon technological advancements in the domain of posture recognition. An innovative Y_PN-MSSD [14] model synergistically combines Pose-Net and Mobile-Net SSD, leveraging the TensorFlow Lite MoveNet framework for enhanced yoga posture recognition. What sets the Y_PN-MSSD model apart is its capability to live-track and correct yoga poses in real-time, demonstrating high accuracy and efficiency. This model is particularly beneficial in online yoga practice environments, where direct instructor guidance is absent. It bridges the gap between traditional and digital yoga practice, ensuring users maintain correct posture, thereby enhancing the overall safety and effectiveness of their practice.

A vision-based methodology for yoga pose grading has been proposed in [15]. The approach centers around the extraction and encoding of human body skeleton key points from images of yoga poses, employing contrastive skeleton feature representations. This method stands out in its ability to handle the inherent variations in pose images, thanks to the use of both coarse and fine triplet examples for feature comparison.

The research gaps identified are listed below:

- While short-term improvements may be observable, tracking progress over an extended period is essential for assessing the efficacy and sustainability of yoga interventions. Existing systems often lack a robust progress monitoring component, making it challenging to evaluate the long-term impact of yoga practice on conditions.
- Progress monitoring in yoga practice is another research area that can be focused. This entails systematic tracking of improvements in accuracy of pose execution, endurance during poses, duration held over time, and overall performance enhancement. Implementing progress monitoring mechanisms could provide valuable insights into the effectiveness of yoga interventions in managing their condition.

The proposed system is developed to address the above gaps by providing real-time feedback and correction on yoga poses. Also, the proposed system takes into consideration the endurance level which is not considered in [14] and includes real time feedback on the fly which is not available in [12]. In addition, the posture retention duration is also measured in the proposed work when compared with [12] and [14]. The proposed work aims to reduce the risk of yoga-related injuries, particularly in populations more prone to such injuries, like older adults. By successfully implementing and testing this system, the work aims to contribute valuable insights and methodologies to the broader field of AI applications in wellness, rehabilitation, and smart health care systems.

Proposed system

Dataset

This work employs a specialized dataset comprised solely of images annotated with yoga posture labels, specifically focusing on poses beneficial for individuals managing spondylitis. This dataset has been compiled from various Kaggle repositories and is significantly augmented by the images provided by paper [16]. The images of 10 different yoga posture namely: Adho Mukha Svanasana (Downward facing dog), Anjaneyasana (Crescent Lunge), Ardha Matsyendrasana (Sitting Half-Spinal Twist), Bhujangasana (Cobra Pose), Dhanurasana (Bow Pose), Marjariasana (Cat Pose), Setu Bandha Sarvangasana (Bridge Pose), Shishuasana or balasana (Child pose), Tadasana (Mountain pose) have been gathered. The images in our dataset have undergone meticulous curation and cleaning to ensure the exclusion of any outliers, guaranteeing a high-quality, consistent set of data for analysis. Table 1 lists the details pertaining to the classwise image count. To handle dataset bias effectively, this approach focuses on joint angles rather than absolute positions. Regardless of body type, the angles between joints for a correctly performed yoga pose are largely consistent, making the evaluation less dependent on individual physical characteristics. This feature helps the model generalize better across users with diverse body types.

The dataset's exclusive focus on yoga asanas related to spondylitis management ensures its direct applicability to our study's goal of enhancing therapeutic yoga practice. By manually extracting and analyzing body landmarks and geometric relationships within these images, our research aims to develop a nuanced understanding of pose execution, offering a novel approach to assessing and improving yoga practice for spondylitis relief, thereby contributing valuable insights to the intersecting fields of health, wellness, and technology.

Proposed SpY_MARNet model

The primary objective of SpY_MARNet model is to revolutionize the way individuals engage with yoga practices, particularly those managing conditions like spondylitis, through the integration of advanced artificial intelligence. This ambitious goal is realized through a meticulously designed system that incorporates several key features and capabilities: (i) a versatile and robust computational framework capable of processing complex body landmark data, ensuring scalability for a wide range of yoga postures; (ii) a user-centric design philosophy that prioritizes intuitive interaction and personalized feedback for practitioners at varying levels of proficiency; (iii) the capacity to efficiently process, analyze, and utilize extensive datasets of body landmarks, angles, and distances, enabling users to receive real-time feedback and insights tailored to their unique practice needs; (iv) sophisticated deep learning algorithms that not only provide deep analytical insights into posture execution but also enhance user interaction by suggesting corrections in real-time. Utilizing color-coded line visuals, the system intuitively indicates the correctness of a pose, with varying colors. This interactive feedback mechanism facilitates a more informed and effective yoga practice, enabling users to adjust their poses instantly and accurately based on the visual cues provided; and (v) the deployment of an innovative artificial dense neural network architecture, designed to seamlessly integrate with physical practice, providing a bridge between traditional yoga and the cutting-edge realm of artificial intelligence. This approach not only enhances the practice of yoga but also contributes significantly to the health and wellness of individuals by offering a smart, responsive, and scientifically grounded practice tool.

Table 1 Dataset class labels and their respective image count

Yoga pose (Class Label)	Image count
Adho Mukha Svanasana (Downward facing dog)	147
Bhujangasana (Cobra Pose)	142
Ardha Matsyendrasana (Sitting Half-Spinal Twist)	153
Bitilasana (Cow pose)	152
Setu Bandha Sarvangasana (Bridge Pose)	155
Anjaneyasana (Crescent Lunge)	151
Tadasana (Mountain pose)	151
Marjariasana (Cat Pose)	140
Shishuasana or balasana (Child pose)	141
Dhanurasana (Bow Pose)	140

Mediapipe

Mediapipe is used in computer vision based applications, where it excels in tasks ranging from facial detection to complex pose estimation. It has 33 extracted landmarks using Mediapipe posture landmarker model. MediaPipe's architecture is designed around a graph-based framework, facilitating the seamless integration and coordination of various processing components. These components, or nodes, are adept at performing specific functions on the streaming data, from initial acquisition to final output analysis. At the heart of MediaPipe's capabilities are its pre-trained machine learning models, which are optimized for real-time applications across platforms. These models enable the extraction of meaningful information from visual data such as identifying key points on human hands, faces, and bodies in images and videos as shown in Fig. 1. This extraction is akin to the convolution operation in CNNs but extends beyond to more specialized tasks. This enables the dynamic detection and tracking of objects and gestures in real-time, providing a foundational technology for developing interactive applications that respond intuitively to human movements and behaviors. In essence, MediaPipe democratizes access to advanced computer vision technologies, offering a robust and efficient solution for developing applications that require real-time processing and interpretation of visual data.

SpY_MARNet model uses Mediapipe to extract posture landmark and perform conversion of landmarks into angles for training and testing as shown in Fig. 2. The angle between joints is computed by considering three consecutive landmarks, typically representing key anatomical points such as the knee, hip, shoulder, and elbow as shown in (1). This computation involves evaluating the arctangent function of the differences in coordinates of these consecutive landmarks.

Let A, B and C be three landmarks, then to find $\angle ABC$:

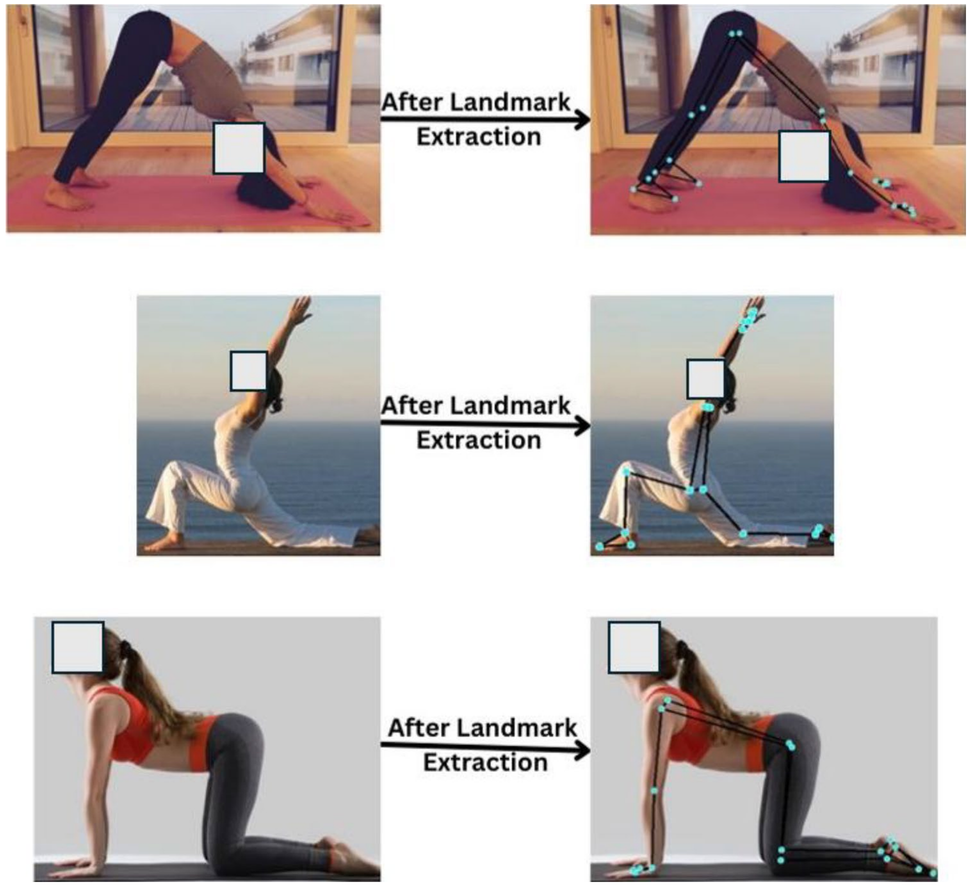

Fig. 1 Mediapipe pose landmarker model tracking 33 body landmarks in real life images

$$\angle ABC, \theta = \left| \tan^{-1}\left(\frac{C_y - B_y}{C_x - B_x}\right) - \tan^{-1}\left(\frac{A_y - B_y}{A_x - B_x}\right) \right| \tag{1}$$

$$if\, \theta > 180 : \theta = 360 - \theta$$

Regularized neural network

A regularized neural network (RegNN) represents an advanced iteration of neural network models, incorporating mechanisms designed to prevent the common problem of overfitting, thereby enhancing the model's ability to generalize to unseen data. Regularization techniques such as Dropout and Batch Normalization play pivotal roles within these networks. Dropout layer, randomly deactivates a subset of neurons during the training process, forcing the network to learn more robust features that are not reliant on any small set of neurons whereas batch normalization standardizes the inputs to a layer for each mini-batch, stabilizing the learning process by reducing the internal covariate shift. Much like the convolution operation in CNNs structures data for hierarchical feature extraction, regularization techniques in such structure combat overfitting, ensuring that the network learns patterns that are truly representative of the underlying data distribution. This makes RegNN exceptionally versatile and effective across a broad spectrum of tasks, from sequence prediction to complex classification challenges, by fostering models that perform well not just on the training data but also on new, unseen data.

SpY_MARNet model (Fig. 3) is trained using a split of 60% for training and 40% for validation, ensuring a robust evaluation of its predictive capabilities. The architecture of the neural network is optimized to capture the nuanced relationships between body postures and their therapeutic efficacy, resulting in a model that is both accurate and reliable. The trained model is then deployed in a real-time environment where it receives landmarks data from users performing yoga poses. This setup allows for immediate analysis and feedback, crucial for ensuring the correct execution of poses and maximizing therapeutic benefits. A reference pose for each of the ten yoga postures is established as a benchmark for evaluating user performance.

Pose accuracy

To assess the accuracy of a user's pose, a novel metric has been used that combines the model's output probability (p1) with a similarity index (p2) calculated using the Cosine Similarity Rule. The Cosine Similarity index (2) compares

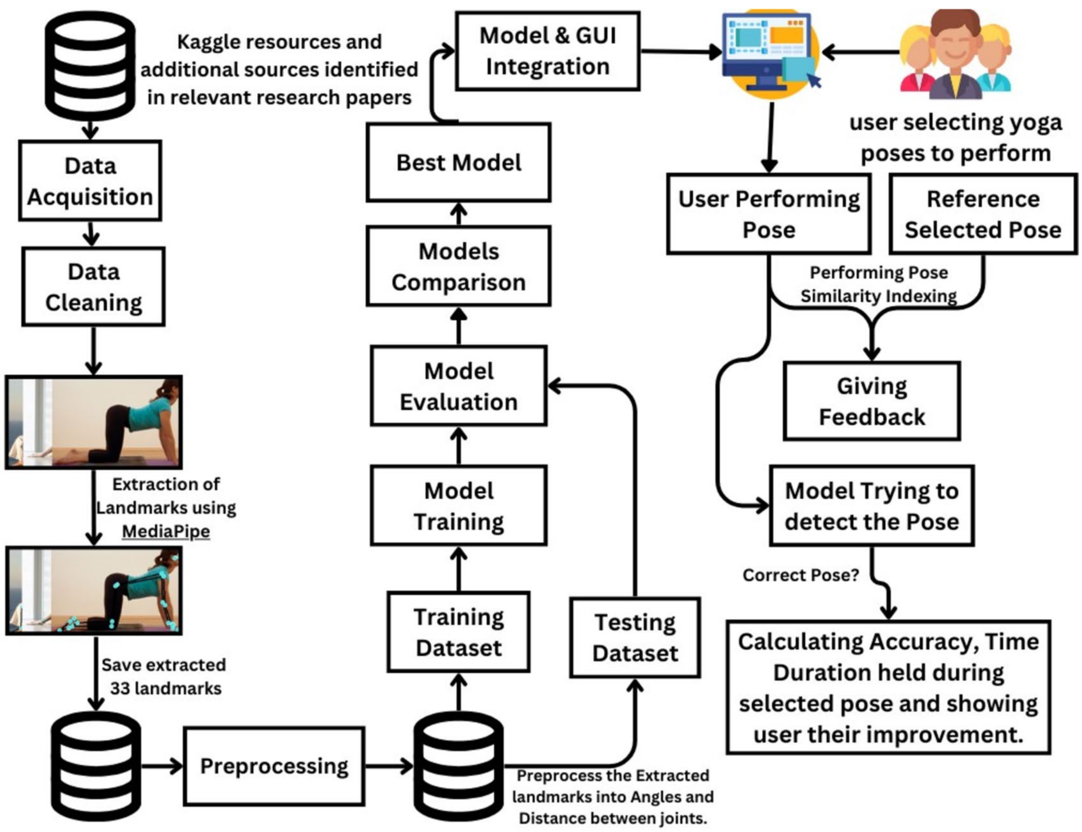

Fig. 2 Flowchart diagram for real-time yoga pose training

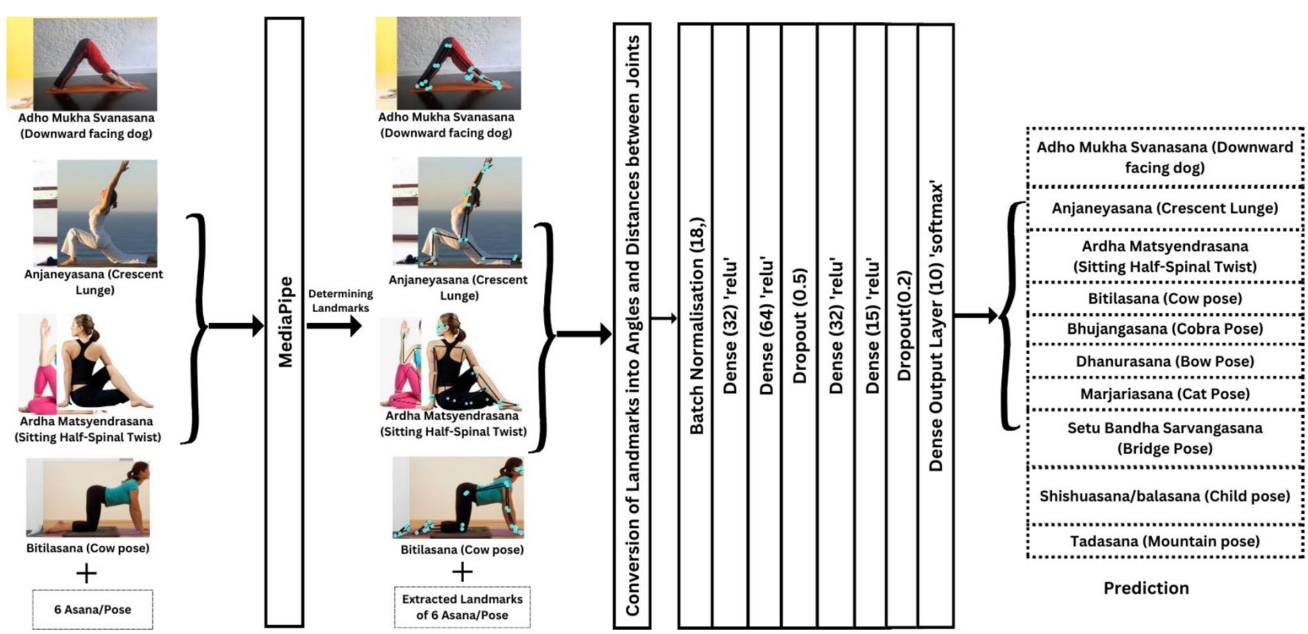

Fig. 3 Architecture of proposed SpY_MARNet model

the user's current pose with a reference pose shown to user represented by the vectors $\overline{A}$ and $\overline{B}$ that denotes angles and distances between different joints using inner product space.

$$\text{similarity, p2} = \left(\frac{A \cdot B}{A \cdot B}\right) * 100 = \left(\frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}\right) * 100 \quad (2)$$

The similarity index (p2) and model's output (p1) are then used to calculate the overall accuracy be doing a weighted sum, where w1 and w2 are weights assigned to p1 and p2 respectively. An equal weight of 0.5 is assigned to both the model's prediction (p1) and the similarity index (p2) in the calculation of overall accuracy (acc) as written in (3) that ensures a balanced evaluation where each aspect are taken into importance equally. By including p1, it ensures that both components are considered equally important in evaluating the user's pose. While p2 measures pose accuracy based on joint angle differences, which is a precise and interpretable metric, p1 represents the model's prediction and captures additional patterns or relationships that may not be directly reflected in p2. Including p1 allows the evaluation to benefit from the model's learned capabilities, such as understanding contextual nuances or higher-order correlations in the data. This combined approach ensures a robust and comprehensive assessment, leveraging the strengths of both components. This approach acknowledges the equal importance of both factors in determining the overall accuracy of the user's pose. By treating both components equally, the evaluation process remains robust, transparent, and flexible, allowing for straightforward adjustments if needed. This simplicity in weighting also enhances user understanding and ensures a fair assessment of performance, promoting a comprehensive evaluation of user poses in the context of yoga practice.

$$\text{acc} = \text{w1} * \text{p1} + \text{w2} * \text{p2} \quad (3)$$

$$\{\text{w1} = \text{w2} = 0.5\}$$

Endurance measurement and user categorization

Endurance plays a very important role in evaluation of a person's ability to sustain a pose for extended durations while maintaining the pose correctly. It is the measure of one's ability to withstand a pose over long duration. It requires mental focus along with muscular engagement throughout the duration of holding the pose, reflecting not only physical but also mental resilience. Incorporating endurance in our work provides a valuable insight into the user's overall fitness along with progress in their yoga journey. The higher the levels of endurance, the greater the strength, flexibility, resilience and focus. By incorporating endurance assessments alongside posture accuracy, a greater understanding of an individual's yoga practice can be obtained to support their overall development and well-being.

This study recognizes that the maximum duration for holding a yoga pose T_{max} may vary depending on factors such as pose complexity, individual fitness levels, and personal preferences. While we utilized a standardized T_{max} of 180 s for our calculations, it is essential to acknowledge that in practical settings, there may not be a fixed or universally applicable maximum duration. As such, our endurance metric, derived from the Eq. (4), may occasionally yield values exceeding 100%. To address this, we have implemented a practical constraint whereby any calculated endurance percentage exceeding 100% is capped at 100%. This approach ensures that our evaluations remain grounded in practicality and accurately reflect participants' endurance levels within the context of their yoga practice, providing meaningful insights for analysis and interpretation.

$$\text{endurance} = \left(\frac{\text{Time}_{\text{duration}} - T_{\text{min}}}{T_{\text{max}} - T_{\text{min}}}\right) * 100 \quad (4)$$

$$\text{if endurance} > 100 : \ \text{endurance} = 100$$

where $T_{min} = 0$ and $T_{max} = 180$ is the minimum and maximum duration to hold a pose respectively to normalize time duration on a scale of 0 to 100%.

Combined metric

This paper's primary goal is to ensure that users are performing yoga poses with a high level of accuracy, but it also value endurance as an important secondary factor. Based on this, it prioritized accuracy more heavily by assigning it a weight of $W_{acc} = 0.7$ and gave a lower weight to endurance, $W_{end} = 0.3$. This reflects the importance of achieving the intended form and alignment in each pose, which is fundamental to reaping the physical and mental benefits of yoga practice.

While accuracy is the main focus, this model still value endurance as it reflects the user's ability to hold poses for extended periods, indicating physical fitness and stamina. By assigning a weight of 0.3 to endurance as shown in Eq. (5), to acknowledge its significance as a secondary factor in overall performance evaluation. The chosen weights strike a balance between accuracy and endurance, giving greater emphasis to accuracy while still considering endurance.

This weighting scheme ensures that both factors contribute meaningfully to the evaluation, aligning with our goal of promoting correct pose execution while also recognizing the importance of physical endurance in yoga practice.

$$\text{present}_{\text{combined}_{\text{metric}}} = W_{acc}*acc + W_{end}*endurance \tag{5}$$

Improvement

This research aimed to quantify the improvement in users' performance metrics over the period of yoga practice. To achieve this, a dynamic calculation of improvement is implemented considering various time frames. A calculation is done as shown in Eq. (6) to compute the average Combined_Metric values for the last 7 days, the 7 days before that, and the rest of the days. In instances where data for the last 7 days or the 7 days before that was unavailable, an adjustment in the weights proportionally is done to maintain accuracy. Here, na denotes not applicable denoting that the data is unavailable.

$$\text{prev}_{\text{combined}_{\text{metric}}} = 0.5*\text{avg}_{\text{last7days}} + 0.3*\text{avg}_{\text{last7days}_{\text{before}}} + 0.2*\text{avg}_{\text{restDays}} \tag{6}$$

$$\text{prev}_{\text{combined}_{\text{metric}}} = 0.6*\text{avg}_{\text{last7days}} + 0.4*\text{avg}_{\text{last7days}_{\text{before}}}, \text{ if } \text{avg}_{\text{restDays}} = \text{na}$$

$$\text{prev}_{\text{combined}_{\text{metric}}} = \text{avg}_{\text{last7days}}, \text{ if } \text{avg}_{\text{restDays}} = \text{na and } \text{avg}_{\text{last7days}_{\text{before}}} = \text{na}$$

Finally, a computation for the improvement as written in Eq. (7) is being done for the user by comparing $\text{present}_{\text{combined}_{\text{metric}}}$ to the $\text{prev}_{\text{combined}_{\text{metric}}}$.

$$\text{improvment} = \frac{\left(\text{present}_{\text{combined}_{\text{metric}}} - \text{prev}_{\text{combined}_{\text{metric}}}\right)}{\text{prev}_{\text{combined}_{\text{metric}}}} * 100 \tag{7}$$

This approach ensured robustness in assessing users' progress despite varying data availability, providing valuable insights into their performance trends for analysis.

Real-time pose estimation interface

This research paper introduced a real-time yoga pose recognition system utilizing pose estimation techniques. The system is designed to analyze video input, which is processed into individual frames for pose estimation. In our pursuit of enhancing model accuracy, an additional step has been incorporated into the system's workflow. Specifically, we predict the pose for six consecutive frames and subsequently employ a majority voting approach to determine the final pose prediction. This strategy acknowledges the inherent imperfections within our model, as no model achieves perfect accuracy. By aggregating predictions across multiple frames and selecting the most commonly occurring pose, we aim to mitigate the impact of occasional inaccuracies, thereby ensuring more robust and reliable predictions. This measure not only enhances the system's overall accuracy but also bolsters confidence in the correctness of the final pose estimation, ultimately improving the user experience and the system's efficacy in supporting yoga practice.

Algorithm Pose Precision: Perfecting Your Chosen Asana {Ps}

Input: Real Time Video

Step 1: Feed the pre-recorded video or live video and convert it into N frames.

Step 2: Utilize MediaPie pose landmarker model to extract 33 body landmark locations.

Step 3: Convert those landmark locations to find angles and distances between joints.

Step 4: Use proposed neural network to detect pose for 6 consecutive frames.

Step 5: Take majority of detected pose for 6 frames $\{P_d\}$.

- If $\{P_d\} = \{P_s\}$, then Yoga pose is recognized and the timer starts. Along with-it, lines connecting the joints changes to green.
- If $\{P_d\} \neq \{P_s\}$, then Yoga pose is not matched, and the connected lines remains red with a proper feedback.
- The user adjusts their pose according to the reference pose untill the pose is detected by out model.

Step 6: After Pose is performed save the Pose Accuracy, Endurance and Combined Metric.

Step 7: Stop the execution.

SpY_MARNet model extracts 33 landmarks from each frame captured by the user's camera. These landmarks serve as pivotal points in inferring the user's posture. Subsequently, a conversion of these landmarks into angles and distances between joints is done, constituting the input data for the regularized neural network. The model then predicts the user's pose based on this input. If the predicted pose matches the selected yoga pose, the connecting lines between the landmarks transition to green as shown in (Figs. 4, 5), signaling correct pose recognition, and initiating the timer.

Conversely, if a mismatch occurs, indicating incorrect pose recognition, the lines remain red, prompting the user to adjust their posture accordingly. This iterative process continues until alignment with the selected pose is achieved. Furthermore, our system provides real-time feedback (Fig. 6) on pose correctness, aiding users in refining their yoga practice with precision.

Additionally, the users can track their progress over time by accessing visualizations such as line and bar charts (Figs. 7, 8). These visual aids showcase metrics including endurance, pose accuracy, combined metric, and improvement over consecutive practice sessions. This feature empowers users to monitor their development and tailor their practice regimen accordingly, fostering continuous improvement and mastery of various yoga poses.

Experimental results and analysis

The comparative performance Table 2 illustrates the efficacy of the Proposed SpY_MARNet in relation to a model utilizing MediaPipe Landmarker features processed through a Random Forest classifier. As evidenced by the metrics, SpY_MARNet model exhibits a superior performance across all evaluated criteria. It achieves an impressive accuracy of

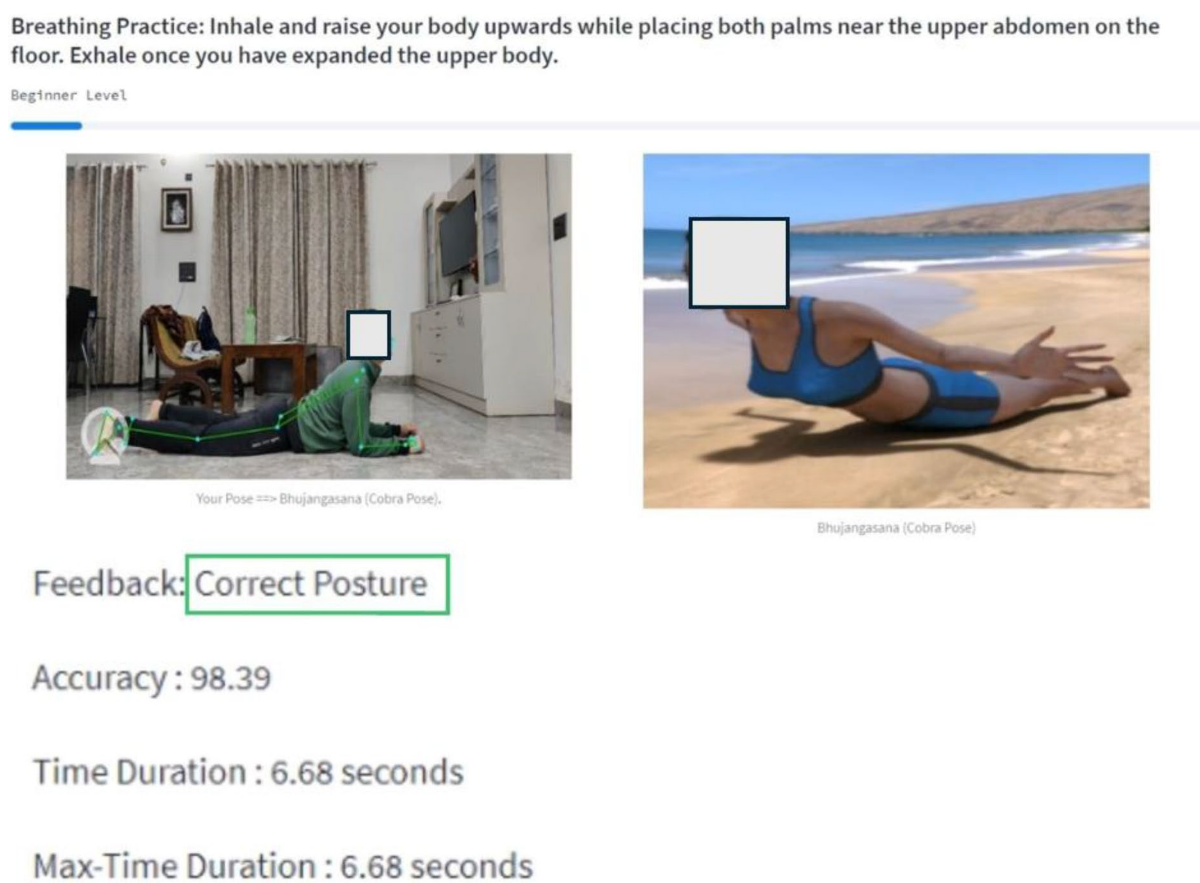

Fig. 4 Half-cobra pose recognized by SpY_MARNet model

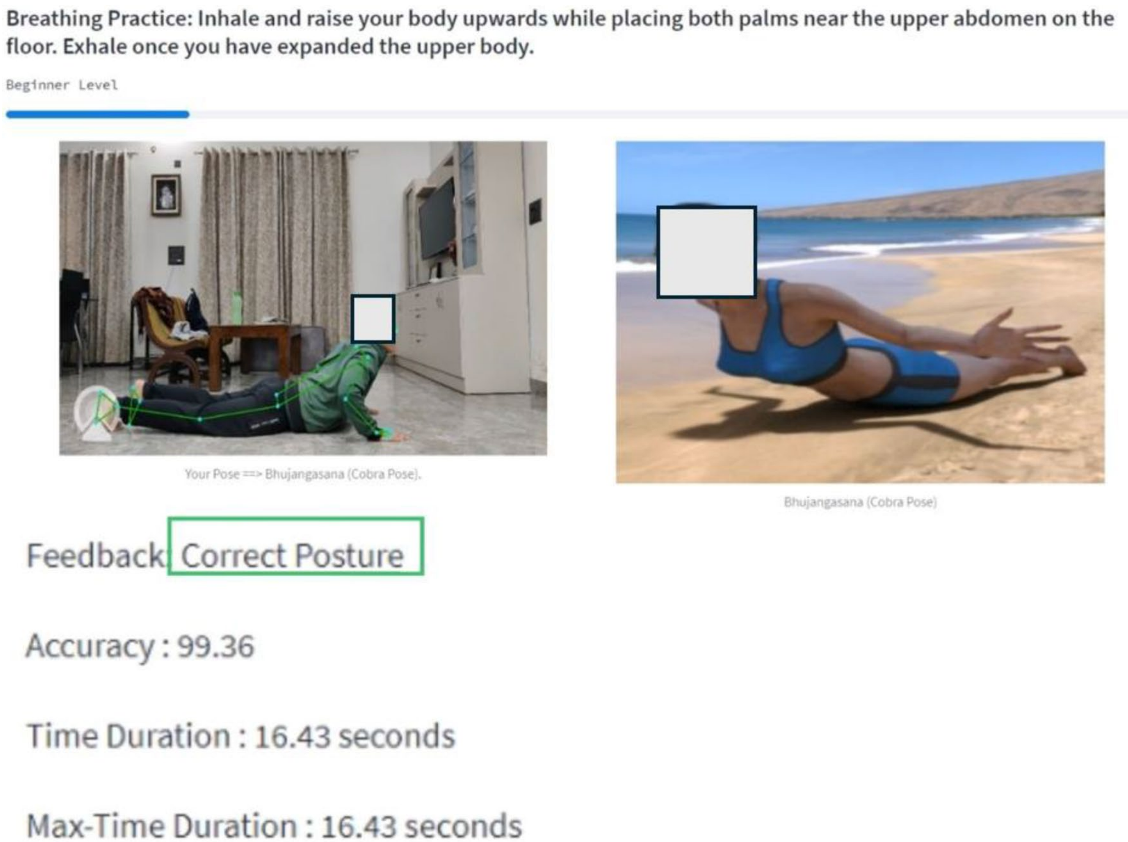

Fig. 5 Cobra pose recognized by SpY_MARNet Model

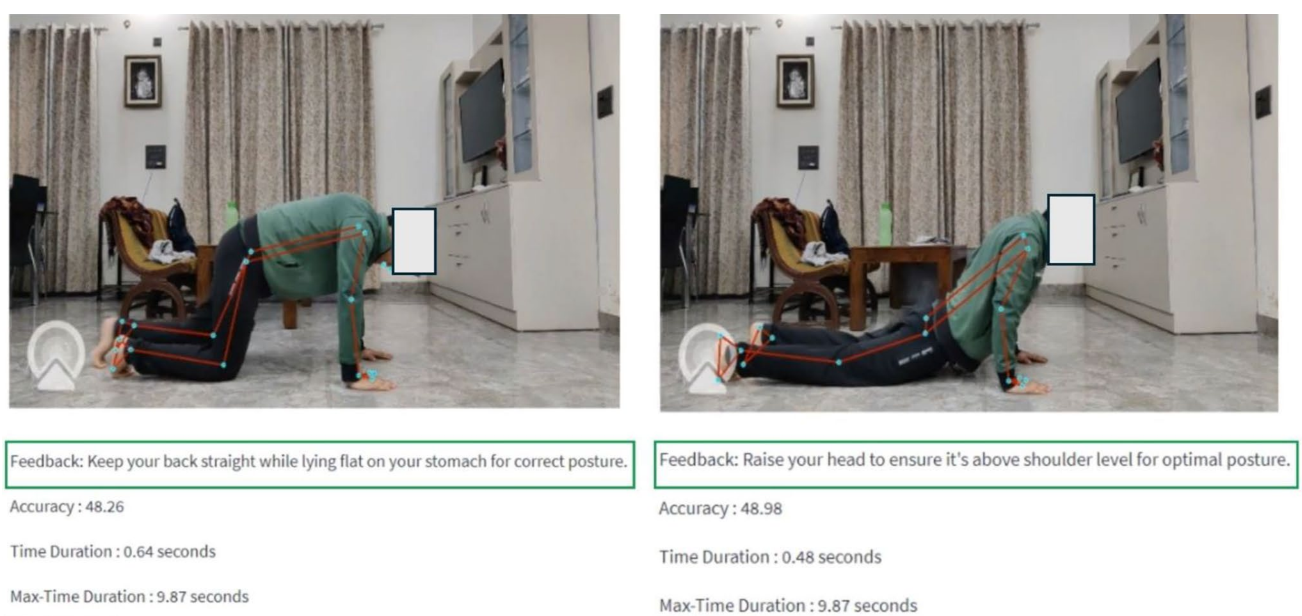

Fig. 6 Real-time feedback given when Cobra pose is not performed the correct way

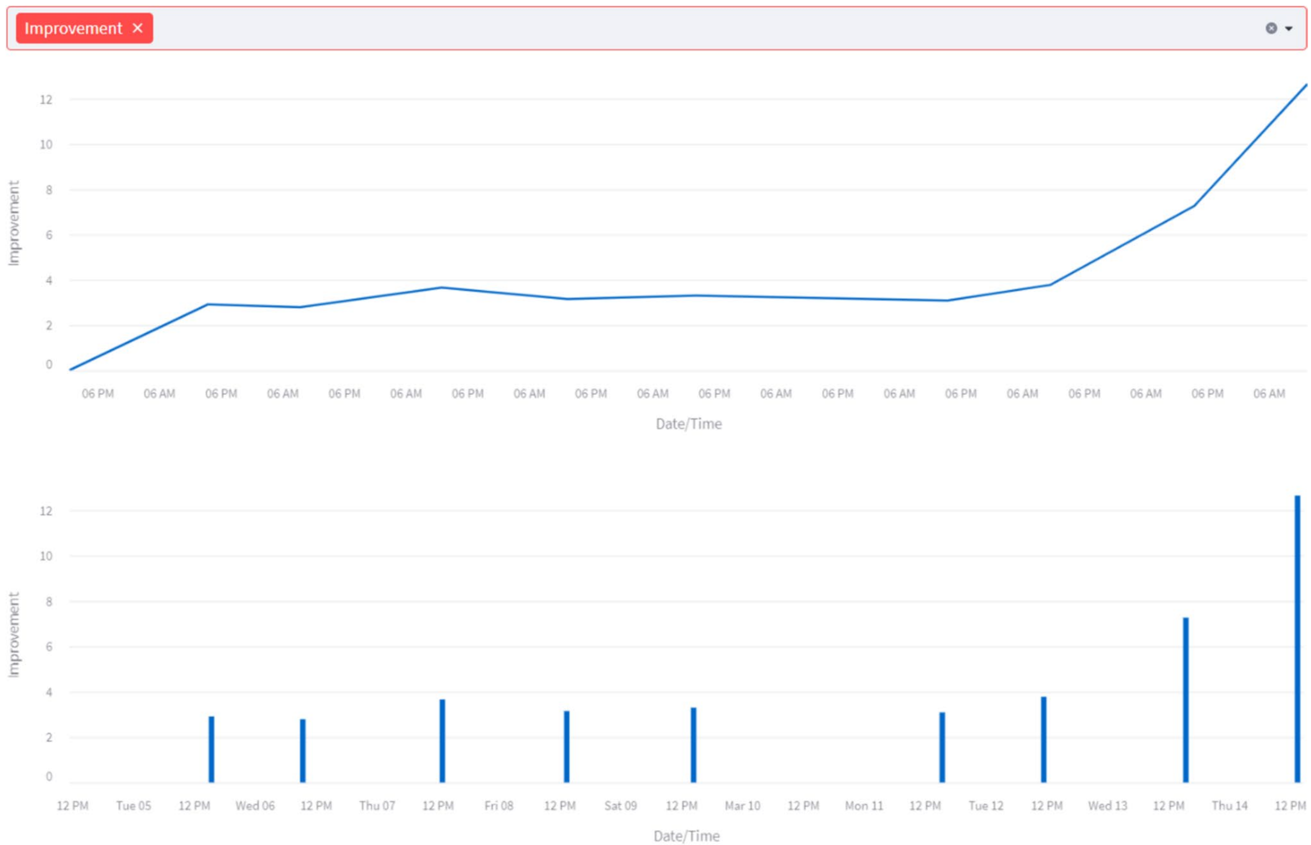

Fig. 7 Graphs denoting user's improvement over the days

99.7%, which is a marginal but noteworthy improvement over the 98.9% accuracy of the MediaPipe Landmarker Random Forest model. Precision, a measure of the model's correctness in identifying a yoga pose as spondylitis-related, is recorded at 99.6% for SpY_MARNet, indicating a higher reliability in pose classification compared to the 98.9% precision of the other model. Similarly, recall, which reflects the model's ability to identify all relevant instances of the spondylitis-related poses, is at 99.5% for SpY_MARNet. These results collectively underscore the robustness of SpY_MARNet in classifying spondylitis-related yoga poses with high precision and recall, demonstrating its potential as a reliable tool for aiding individuals with spondylitis in their yoga practice.

In this research, an extensive training and validation of our model over 1000 epochs is conducted to evaluate its performance and stability. Throughout this iterative process, we meticulously tracked both training and validation accuracy to understand how well our model generalized to unseen data. Our analysis revealed intriguing insights, notably observing a clear trajectory of improvement in accuracy over epochs. Strikingly, we identified the epoch 615 as pivotal (Fig. 9), where the model achieved its peak accuracy. This convergence at epoch 615 underscores a significant milestone in our research where we have saved our model, signifying a delicate balance between model complexity and generalization capability. The training and validation loss plot over the epoch is provided in (Fig. 10).

The confusion matrix in Fig. 11 illustrates the performance of a machine learning model in classifying yoga poses from images. The rows represent the ground truth labels, and the columns represent the predicted labels. The diagonal elements represent the number of correctly classified instances for each pose. For instance, the model perfectly classified all 147 downward-facing dog (Adho Mukha Svanasana) poses, and none were misclassified as other

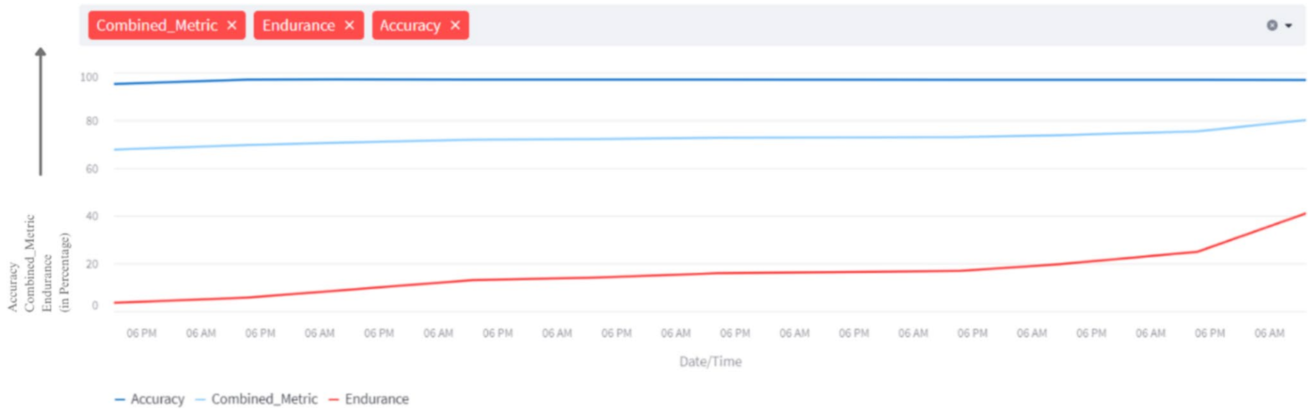

Fig. 8 Graph depicting the user accuracy, endurance and combined_metric details over a duration

Table 2 Performance metrics of proposed SpY_MARNet model

Model	Mediapipe landmarker with random forest	Proposed SpY_MARNet
Accuracy	98.9%	99.7%
Precision	98.9%	99.6%
Recall	98.9%	99.5%

poses. Overall, the model achieved a classification accuracy of 696 out of 698 poses.

Table 3 presents a comprehensive evaluation of the proposed SpY_MARNet model using a variety of optimization algorithms to determine their impact on classification accuracy. The Adam optimizer, which is the algorithm of choice within the SpY_MARNet framework, demonstrates superior performance with a maximum accuracy of 99.7%. This is significantly higher compared to the other optimizers tested. Stochastic Gradient Descent (SGD) yields a modest accuracy of 80.9%, while RMSprop exhibits a competitive performance with an accuracy of 99.14%. Interestingly, the Adadelta and Adagrad optimizers show considerably lower accuracies of 9.1% and 32.09% respectively, suggesting they may not be well-suited to the model or the specificities of the task at hand. These findings highlight the effectiveness of the Adam optimizer in the context of our network and underscore the importance of optimizer selection in the performance of deep learning models.

The proposed model offers users the functionality to select a specific yoga pose and attempt to replicate it based

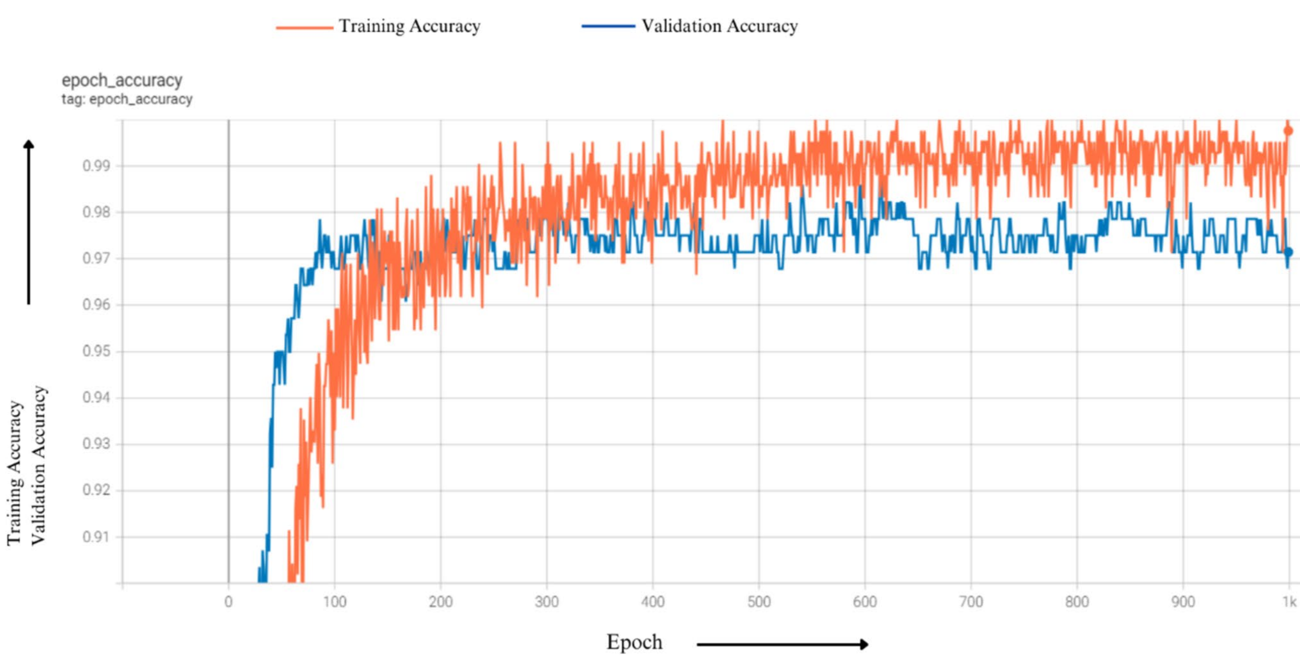

Fig. 9 Training and validation accuracy v/s epochs graph

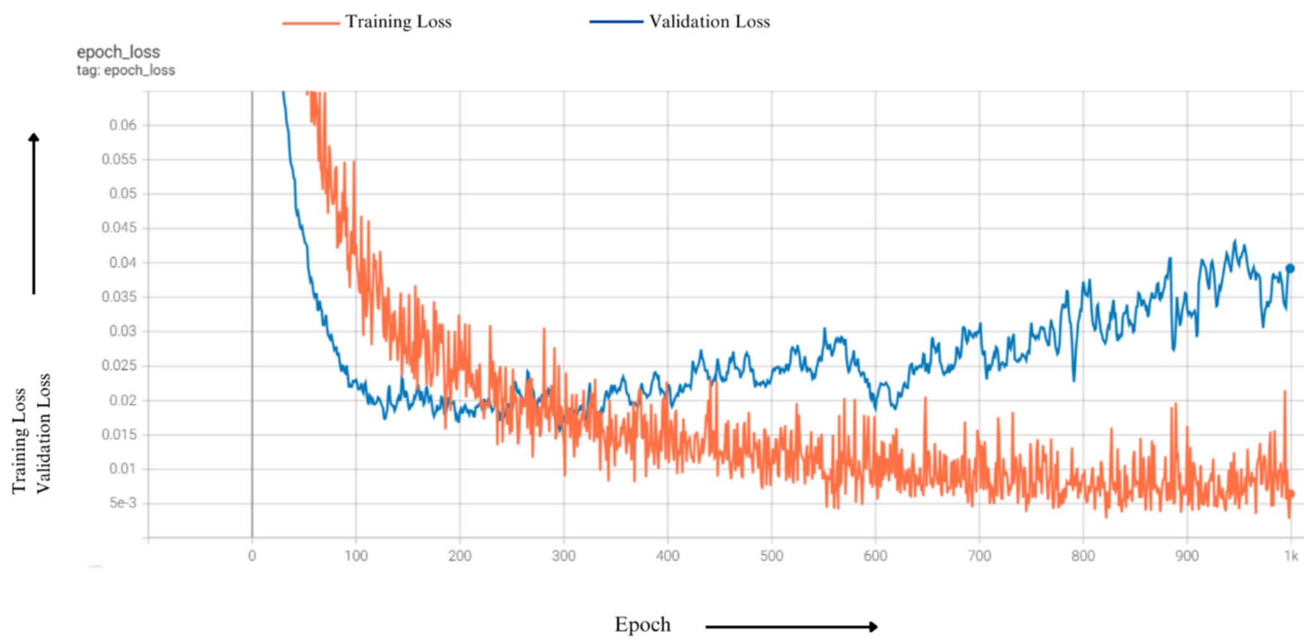

Fig. 10 Training and validation loss v/s epoch's graph

on a reference image until their pose is accurately recognized by our model. Upon successful recognition of the posture, an auditory cue signals the commencement of a timer, initiating concurrent calculations for accuracy and endurance metrics until the pose is no longer recognized. All pertinent data, including accuracy measurements and endurance levels as highlighted in Figs. 12 and 13. For Bridge Pose and Downward Facing Dog Pose respectively are meticulously recorded for future reference by the user. This systematic approach ensures a comprehensive evaluation of the user's performance while engaging in yoga practice, facilitating informed progress tracking and personalized refinement of their technique over time.

Conclusion and future work

The proposed SpY_MARNet leverages the power of artificial intelligence to bring a new dimension to yoga practice, offering a personalized, accurate, and therapeutically beneficial experience to users. By integrating real-time pose analysis, accuracy measurement, and endurance tracking, we aim to contribute significantly to the fields of health, wellness, and AI-driven personal fitness, particularly for those managing conditions such as spondylitis. This model has flexibility for adjustment as these weights are not set in stone and can be adjusted based on user feedback, project objectives, or domain expertise. They provide a starting point for evaluation, and we can iterate and refine our approach as needed to better reflect the priorities of our target audience and the goals of our work. Also, the proposed SpY_MARNet provided an accuracy of 99.7%. The data collected of user performing various yoga poses, including accuracy, time taken to complete each pose, and endurance levels, serves as a valuable resource for further research. By analyzing the metrics generated by the user during their yoga journey, researchers can gain deeper insights into the physiological aspects of yoga practice. It can help to clarify how accuracy and endurance can together shed light on the effectiveness of different poses in promoting physical fitness and mental well-being. Additionally, such data can inform the development of personalized yoga routines tailored to individual needs and goals, enhancing the overall effectiveness and accessibility of yoga practice for spondylitis. By understanding yoga's benefits more comprehensively, this information can inform strategies for improving physical health, mental wellness, and performance. The model has certain limitation in that it is trained on yoga poses specific to Spondylitis treatment. This might restrict its applicability for individuals practicing a broader range of yoga exercises. For this, an expansion of the dataset is required to include more diverse poses and conditions to make the model more

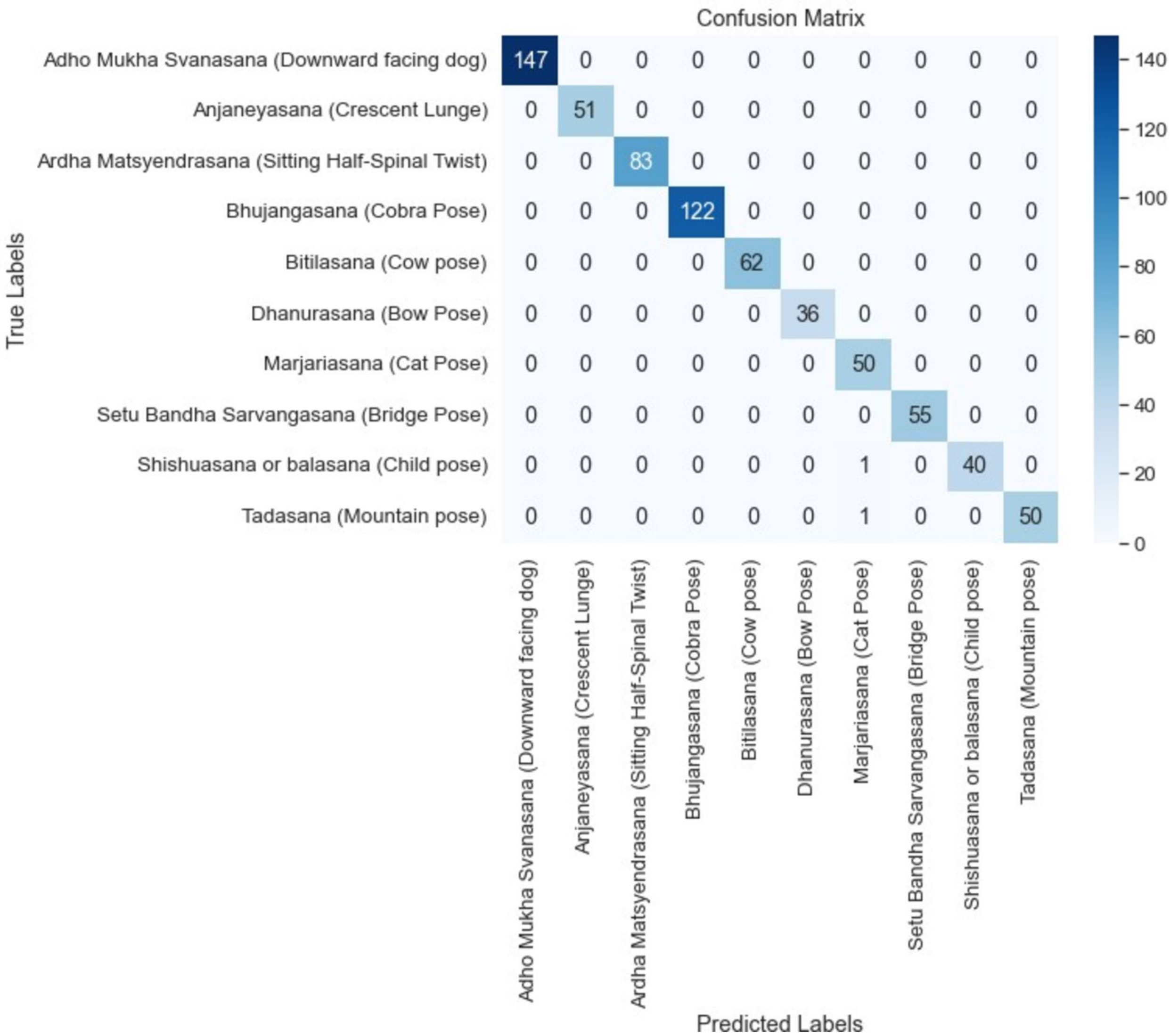

Fig. 11 Confusion matrix of SpY_MARNet model

Table 3 Performance metrics of proposed SpY_MARNet model using different optimizers

S. No	Optimizer	Maximum accuracy obtained in percentage
1	Adam (used by SpY_MARNet)	99.7
2	SGD	80.9
3	RMSprop	99.14
4	Adadelta	90.1
5	Adagrad	32.09

versatile and train the model better. Another limitation is that the model focuses on static pose evaluation, which might not capture the dynamics of yoga poses involving transitions or movements. To capture such poses model needs to be extended to analyze sequences of poses or transitions using models like LSTMs or transformers to evaluate the fluidity and accuracy of movements. This can be a potential enhancement in future. Future scope can be towards scaling up the model by training this model to understand more yoga poses for other health problems and to include the dynamic pose evaluation as well.

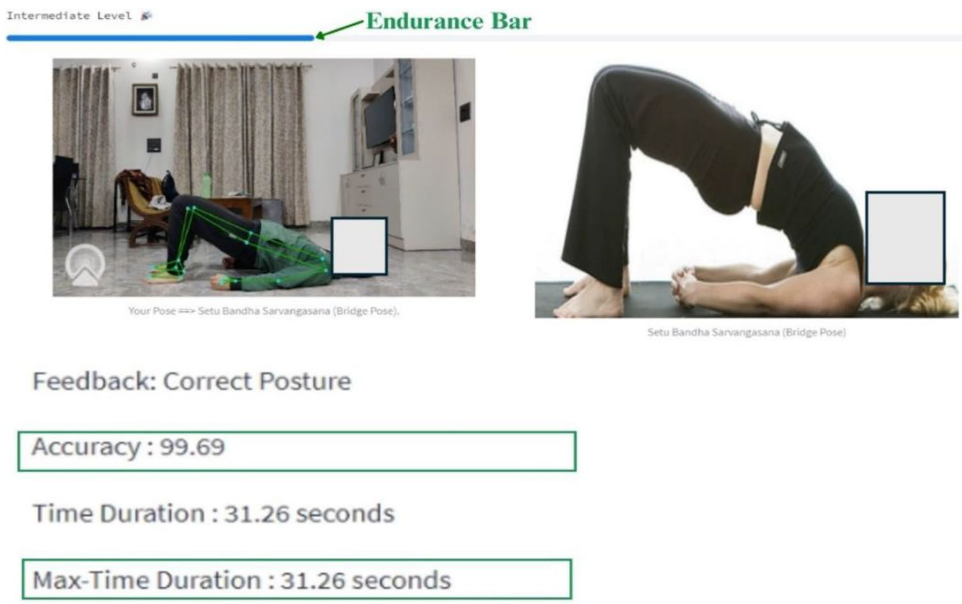

Fig. 12 Illustration of correct bridge pose with endurance bar

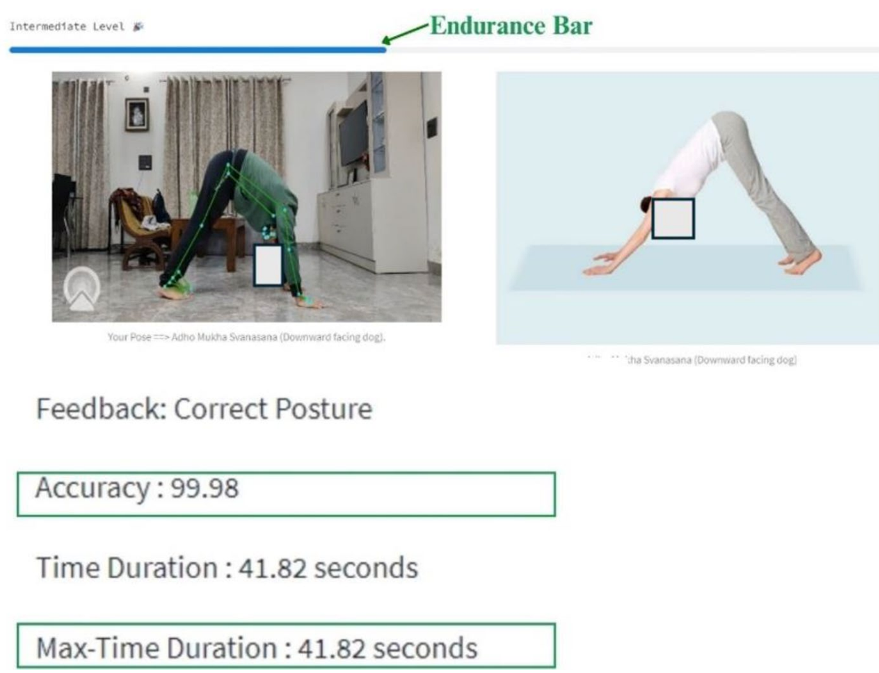

Fig. 13 Illustration of correct downward facing dog pose with endurance range bar

Acknowledgements We would like to thank VIT Chennai management for the support and encouragement provided to pursue this work.

Author contributions A.R. and A.B. did the ideation, conceptualization and verification. A.R., A.B and C.A.V did the implementation and manuscript writing. All authors reviewed the manuscript.

Funding This work did not involve any funding.

Data availability The data used for this work was taken from the repository available in the link: https://data.mendeley.com/datasets/jc4mmnvcdk/1.

Declarations

Conflict of interest The authors declare no competing interests. All the authors declare that there are no potential conflict of interest.

References

1. Swain TA, McGwin G (2016) Yoga-related injuries in the United States from 2001 to 2014. Orthop J Sports Med 4(11):2325967116671703

2. Bulat A., & Tzimiropoulos G. (2016) Human pose estimation via convolutional part heatmap regression. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14 (pp. 717–732). Springer International Publishing.
3. Sharma A., Shah Y., Agrawal Y., & Jain P. (2022) Real-time recognition of yoga poses using computer vision for smart health care. arXiv preprint arXiv:2201.07594.
4. Kendall A., Grimes M., & Cipolla R. (2015) Posenet: A convolutional network for real-time 6-dof camera relocalization. In Proceedings of the IEEE international conference on computer vision pp. 2938–2946
5. Hsieh C. C., Liou D. H., & Lee D. (2010) A real time hand gesture recognition system using motion history image. In 2010 2nd international conference on signal processing systems (Vol. 2, pp. V2–394). IEEE
6. Kodama T., Koyama T., & Saitoh T. (2017) Kinect sensor based sign language word recognition by mutli-stream HMM. In 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE) (pp. 94–99). IEEE.
7. Agrawal Y., Shah Y., & Sharma A. (2020) Implementation of machine learning technique for identification of yoga poses. In 2020 IEEE 9th international conference on communication systems and network technologies (CSNT) (pp. 40–43). Ieee.
8. Lei Q, Du JX, Zhang HB, Ye S, Chen DS (2019) A survey of vision-based human action evaluation methods. Sensors 19(19):4129
9. Li J, Hu Q, Guo T, Wang S, Shen Y (2021) What and how well you exercised? An efficient analysis framework for fitness actions. J Vis Commun Image Represent 80:103304
10 Bukhari M, Bajwa KB, Gillani S, Maqsood M, Durrani MY, Mehmood I, Rho S (2020) An efficient gait recognition method for known and unknown covariate conditions. IEEE Access 9:6465–6477
11. Bazarevsky V., Grishchenko I., & Raveendran K. (2006) BlazePose: On-device Real-time Body Pose tracking. 2020. 10.48550. ARXIV
12. Islam M. U., Mahmud H., Ashraf F. B., Hossain I., & Hasan M. K. (2017) Yoga posture recognition by detecting human joint points in real time using microsoft kinect. In 2017 IEEE Region 10 humanitarian technology conference (R10-HTC) (pp. 668–673). IEEE
13. Chu X., Ouyang W., Li H., & Wang X. (2016) Structured feature learning for pose estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 4715–4723
14. Upadhyay A, Basha NK, Ananthakrishnan B (2023) Deep learning-based yoga posture recognition using the Y_PN-MSSD model for yoga practitioners. In Healthcare 11(4):609
15. Wu Y, Lin Q, Yang M, Liu J, Tian J, Kapil D, Vanderbloemen L (2021) A computer vision-based yoga pose grading approach using contrastive skeleton feature representations. In Healthcare 10(1):36
16. Suryawanshi Y, Gunjal N, Kanorewala B, Patil K (2023) Yoga dataset: A resource for computer vision-based analysis of Yoga asanas. Data Brief 48:109257
17. Wang J, Yuan S, Tuantuan Lu, Zhao H, Zhao Y (2024) Fusing YOLOv5s-MediaPipe-HRV to classify engagement in E-learning: from the perspective of external observations and internal factors. Knowl-Based Syst 305:112670