DeepYoga: Enhancing Practice with a Real-Time Yoga Pose Recognition System

Roli Bansal
Department of Computer Science, Keshav Mahavidyalaya, University of Delhi, India
roli.bansal@keshav.du.ac.in

Richa Sharma
Department of Computer Science, Keshav Mahavidyalaya, University of Delhi, India
richasharma@keshav.du.ac.in (corresponding author)

Priyanshi Jain
Department of Computer Science, Keshav Mahavidyalaya, University of Delhi, India
priyanshi2216120@keshav.du.ac.in

Rahul Arora
Department of Computer Science, Keshav Mahavidyalaya, University of Delhi, India
rahul2216063@keshav.du.ac.in

Sourabh Pal
Department of Computer Science, Keshav Mahavidyalaya, University of Delhi, India
sourabh2216110@keshav.du.ac.in

Vishal
Department of Computer Science, Keshav Mahavidyalaya, University of Delhi, India
vishal2216109@keshav.du.ac.in

Received: 6 August 2024 | Revised: 3 September 2024 and 14 September 2024 | Accepted: 16 September 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.8643*

ABSTRACT

The adoption of yoga as a holistic wellness practice is increasing throughout the world. However, in the absence of a personalized expert, especially in an online environment, there is a need for reliable and accurate methods for yoga posture recognition. Maintaining correct yoga postures is essential to reap holistic health benefits in the long term and address chronic medical issues. This paper presents DeepYoga, a novel approach to improve posture recognition accuracy with the support of deep learning models. The proposed approach uses a dataset of accurate yoga pose images encompassing five distinct poses. Landmarks extracted from the practitioner's body in the images are then used to train a Convolutional Neural Network (CNN) for accurate pose classification. The trained model is then used to detect yoga poses from real-time videos of yoga practitioners. Then, the system provides users with real-time feedback and visual suggestions, helping them improve physical alignment and reduce the risk of injury. The proposed method achieved an overall high accuracy of 99.02% in pose detection while trying to minimize the use of resources as much as possible to make it more accessible.

Keywords-deep learning; CNN; real-time yoga pose recognition; personalized feedback; landmarks

I. INTRODUCTION

In today's fast-paced society, people are committed to busy routines, undervaluing the importance of self-care amidst everyday responsibilities, leading to numerous lifestyle health problems. Millions of people worldwide are affected by physical and mental diseases, which highlights the need for efficient methods to support their overall health [1]. Yoga is a traditional Indian form of exercise that addresses the mental, physical, and spiritual well-being of individuals. It is a deep philosophy that aims to unite the mind, body, and soul. Its focus on breath awareness and mindfulness promotes overall well-being by providing the practitioner with skills to reduce stress, increase flexibility, and develop inner peace. Although it originated in Vedic India, yoga has become an internationally recognized discipline known for its numerous benefits, including improved flexibility, muscle strengthening, increased stamina, and general physical and mental health. However, ensuring correct posture during yoga sessions is essential, as incorrect positioning can be ineffective or even harmful. Discrepancies between an instructor's guidance and a user's adopted posture can lead to suboptimal practice. To address this challenge, automated yoga pose detection systems are needed to facilitate accurate and efficient yoga practice, accessible to practitioners of all ages and health conditions. The use of technology has been integrated into daily life due to the increasing use of smartphones and other mobile devices. The latest wearable technology and fitness applications have recently seen a huge increase in popularity, as more and more people rely on them to monitor their exercise patterns and progress [2]. Existing research in this area has explored various technology models for human posture detection, employing techniques such as simulators, sensors, and machine learning algorithms. However, challenges remain in developing systems capable of real-time detection and correction of yoga poses, particularly in diverse self-taught scenarios.

Building on insights from the existing literature, this study aimed to address these challenges by introducing an artificial intelligence-based system for accurate yoga pose recognition. The proposed approach leverages deep learning techniques to develop a framework for automatic yoga pose detection and accuracy evaluation in real-time. Yoga pose detection has been studied using machine learning and deep learning techniques, moving from approaches such as Support Vector Machine (SVM) and Random Forests (RF) to more sophisticated Convolutional Neural Networks (CNN). CNNs belong to the field of deep learning, a subset of artificial intelligence. Their primary advantage over other deep learning algorithms is their minimal preprocessing requirement. CNN-based classification facilitates quick and precise decision-making [3, 4]. A primary objective was to identify gaps in existing methodologies and then highlight opportunities for further advances in the field of yoga pose detection. In [5, 6], an angular approach was used, in which the angles between vectors in a skeletonized image were used to determine and classify poses. However, the proposed model is trained exclusively on image datasets and predicts poses using a landmark-based approach. This facilitates real-time pose prediction without the need for angular calculations, offering a more streamlined and efficient solution. In [7], a real-time yoga pose detection mechanism was proposed to help practitioners correct their posture immediately and practice safely and effectively. This model faced difficulties in identifying the correct yoga poses due to varied conditions such as attire, body shapes, and the complexity of some postures, which can affect its accuracy and reliability. In [8], a yoga posture recognition system was proposed that was able to recognize the pose and collect relevant information from various web-based resources to make users aware of the correct posture. The use of Kinect and star skeleton computation added a layer of complexity that might not be necessary for all applications. The process of body map capturing, contour extraction, and star skeleton computation can be computationally intensive, potentially affecting real-time performance. The model proposed in [9] achieved excellent accuracy on large and diverse yoga datasets. However, it still has some drawbacks, as it may face generalization issues due to physiological variances among users, making it challenging to extrapolate findings to all individuals. The model in [10] shows the correct reference pose alongside the real-time prediction as feedback. However, this approach is not entirely convincing, as it does not effectively guide users on how to improve their posture. This study aims to enhance user guidance by displaying the predicted pose along with a real-time accuracy level. This allows users to adjust and refine their posture more effectively, receiving immediate feedback on how well they are performing.

The accuracy and performance of existing approaches are highly dependent on the quality and diversity of the input images and the datasets used. Although the accuracy of recognizing yoga poses from images or video frames has greatly increased due to these procedures, managing positional variances brought on by attire, background, and lighting, as well as enhancing model generalization across a range of yoga styles and populations, still provides challenges. Motivated by these, the major contributions of the proposed model, called DeepYoga, are:

- Real-time detection: DeepYoga has the potential to recognize yoga poses in real-time, regardless of attire and lighting conditions, providing immediate feedback to practitioners to improve the efficiency and effectiveness of their yoga practice.
- Personalized feedback: By analyzing individual performance, the system provides personalized guidance to help users improve their postures, ensuring a user-specific experience.
- Lightweight architecture: Designed to be lightweight, the model can run efficiently on standard devices without intensive use of resources.

The system has an interactive user interface that facilitates real-time video processing with the help of OpenCV, enabling personalized feedback in real-time to significantly help users in their yoga practice.

II. PROPOSED METHOD

The proposed system, DeepYoga, aims to analyze the particular yoga posture of an individual and provide personalized feedback, ranging from poor to excellent, from a video stream. The system first trains a CNN with precise body landmarks extracted from images of the chosen dataset. The trained model is then used to process the extracted landmarks from video frames in real-time to detect the yoga pose of a user and calculate confidence scores. This provides the user with real-time feedback and visual suggestions to improve his posture.

A. *Workflow*

The proposed system, shown in Figure 1, takes as input either images of individual yoga postures or frames from a video stream in real-time or a video recording. Subsequently, landmarks are detected in each frame to draw the skeletal representation of the posture. These extracted landmarks are stored in a structured format for further processing. These landmarks consist of the *x*, *y*, and *z* coordinates of each key point on the body, providing a comprehensive spatial representation. Then, they are passed to a trained CNN for classification. The accuracy of each classified posture is evaluated as shown in Table I. This real-time feedback mechanism helps users achieve more accurate yoga poses.

B. Dataset

The proposed method was evaluated on the Yoga Poses Dataset [11], which comprises 5 distinct classes, namely (1) Tree, (2) Downdog, (3) Goddess, (4) Plank, and (5) Warrior2. The dataset contains a total of 1551 key pose images for these Yoga classes, separated into training and testing subsets. Figure 2 shows sample images depicting various yoga poses from this dataset. The initial dataset consisted of 1,081 training images and 470 test images. After removing images where landmarks were not detected successfully, the dataset was reduced to 1,026 training images and 457 test images. To enhance the diversity and robustness of the training data, several data augmentation techniques were applied, including rotation, horizontal flipping, and adjustments to brightness and contrast. This augmentation expanded the training set to 2,162 images. Subsequent filtering processes were carried out to ensure data quality, resulting in a final dataset comprising 1,817 training and 428 test images.

TABLE I. POSTURE ACCURACY AND FEEDBACK TO THE USER

Posture accuracy	Feedback
≥ 85%	Excellent
≥ 70% and < 85%,	Good
≥ 40% and < 70%,	Could be better
< 40%	Poor

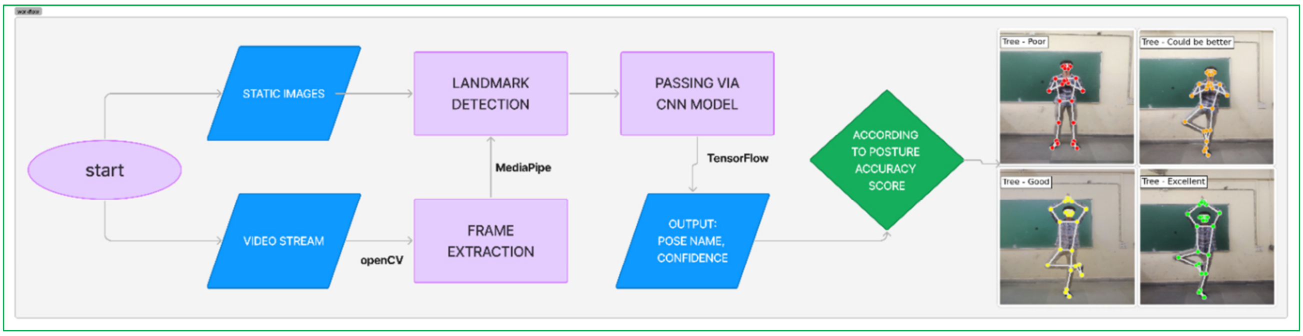


Fig. 1. Workflow of DeepYoga, a real-time yoga pose recognition system.

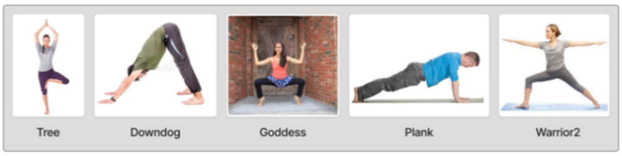


Fig. 2. Yoga postures in the Yoga Poses Dataset [11].

C. Preprocessing

Each image of the dataset was preprocessed using the pose estimation module presented in Figure 3. The module is configured using MediaPipe [12] to identify 33 key landmarks on the human body, covering critical points such as the hips, knees, ankles, shoulders, elbows, and wrists. A constraint during the training and testing phase ensures that if all 33 landmarks are not detected on an image, notwithstanding the attire and lighting conditions, then that image is excluded from the training and testing set to maintain accuracy and optimize model performance. This resulted in a 5% and 2.844% reduction in the total number of images in the training and testing sets, respectively.

D. Model Fitting (Training)

The CNN model designed for yoga pose classification consists of several layers to effectively process the input landmarks of the body pose. Figure 4 shows the CNN architecture for yoga pose detection with personalized feedback. The designed model starts with an input layer for the 3D coordinates of the landmarks, followed by two convolutional layers with ReLU activation and max pooling layers to extract and downsample features. A global average pooling layer reduces the feature maps before passing them to fully connected dense layers with ReLU activation, interspersed with dropout layers to prevent overfitting. The final dense layer uses a softmax activation function to classify the poses into five categories as output. The model utilizes a sparse categorical cross-entropy loss function and Adam optimizer and is trained on the dataset with a validation split of 20% over 120 epochs.

In pose classification, cross-entropy loss measures the difference between predicted probabilities and true class labels using the following equation:

$$L = -\sum_{i=1}^{C}(y_i) \quad (1)$$

where C is the number of classes, y_i is the true label (one-hot encoded), and $\hat{y}_i$ is the predicted probability for class i. This loss penalizes incorrect predictions, with a higher loss indicating greater discrepancy and a lower loss suggesting better alignment. The loss is minimized during training to improve classification accuracy by encouraging higher probabilities for correct poses.

Fig. 3. Pose estimation module.

The softmax function, applied in the final layer, converts raw output scores (logits) into class probabilities. For a vector of logits $z = [z_1, z_2, \ldots, z_C]$, where, C represents the number of classes, the softmax function is as follows:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad (2)$$

This ensures probabilities sum to 1, representing the likelihood of each class. The ReLU function, applied in intermediate layers, introduces non-linearity, defined as:

$$ReLU(x) = \max(0, x) \quad (3)$$

This function outputs x if x is positive, and 0 otherwise. By activating neurons only for positive inputs, $ReLU$ promotes sparsity and computational efficiency in the network.

E. Real-Time Landmark Extraction

To achieve real-time Yoga pose detection and correction, OpenCV [13] or video capture and frame processing were used, following:

- Video capture: Set up a webcam to capture real-time video streams using OpenCV.
- Continuous reading frames: Read the video frames continuously in a loop, so that the system can process them in real-time.
- Frame preprocessing: Convert the frames to the format accepted by MediaPipe for accurate landmarks extraction. Perform required preprocessing steps such as cropping and resizing to prepare the frames for analysis.
- Landmark extraction in real-time: Process each video frame and extract landmarks in real-time.
- Optimization: Optimize the process of landmarks extraction to assure minimal latency and maintain high frame rates, providing a seamless user experience.

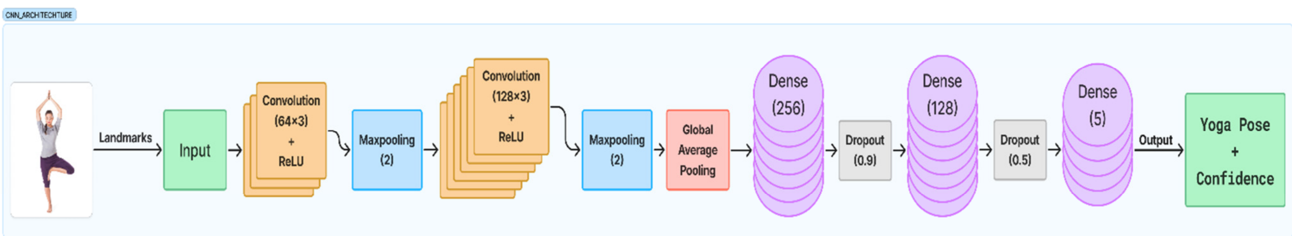


Fig. 4. CNN architecture of DeepYoga with personalized feedback.

F. Pose Detection and Posture Accuracy Calculation

The extracted landmarks from the frames of the real-time video stream are used to detect the yoga pose and calculate the confidence score using the trained CNN model with the following steps:

- Pose detection: The landmarks extracted from the real-time frames are fed to the model to classify the pose. The trained model detects the yoga pose with an associated confidence score, indicating the possibility of the detected pose.
- Posture accuracy: To evaluate the accuracy of a pose, the Euclidean distance is used to determine the similarity between the feature vectors extracted from the input images. After the CNN generates these vectors, the Euclidean distance can quantify the dissimilarity between poses. By incorporating Euclidean distance into the loss function, the model can be optimized to minimize the distance between feature vectors of identical poses and maximize the distance between vectors of distinct poses, thus enhancing the accuracy of pose classification.

  Let $P_k = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_n, y_n, z_n)\}$ be the set of n key points detected. The Euclidean distance (D) between key points in the detected and reference pose can be calculated as:

$$D = \sqrt{\sum_{i=1}^{n}[(x_i - x_i')^2 + (y_i - y_i')^2 + (z_i - z_i')^2]} \quad (4)$$

where, (x_i, y_i, z_i) are key points in the detected pose, and (x_i', y_i', z_i') are key points in the reference pose. This approach helps to overcome the challenges faced by previous models [5, 7, 9], particularly in managing variations in physiological conditions, clothing, and body shapes.

- Real-time feedback: Based on the detected yoga pose and posture accuracy, real-time feedback is provided to the user, which helps to adjust his/her posture as needed.

- User interface: A friendly user interface was implemented to show the detected poses and confidence score so that users can easily understand and improve their poses. This interactive approach increases the user's ability to perform yoga poses safely with high accuracy.

III. RESULTS

The trained model gives a strong performance on the testing dataset in the process of yoga pose detection. With its consistent performance across many indices, the DeepYoga model is a potential tool for automated real-time yoga pose recognition, suggesting robustness and reliability. To evaluate the performance of the proposed model in real-world conditions, real-time images of volunteers performing different yoga poses were analyzed.

An experiment was conducted on a CNN architecture (referred to as Model 1) featuring four dense layers with sizes of 512, 256, 128, and 5 neurons, respectively. Dropout layers with a rate of 0.5 were included after the Dense 512 and Dense 256 layers to prevent overfitting. However, the model still exhibited signs of overfitting, possibly due to the complexity introduced by the excessive number of dense layers. To address this issue, several adjustments were made to the network architecture. The refined architecture (referred to as Model 2) had the following features:

- Removal of layers: The Dense 512 layer and its corresponding dropout layer were removed to reduce model complexity.
- Adjustment of dropout rates: The dropout rate after the Dense 256 layer was increased from 0.5 to 0.9 to further mitigate overfitting by promoting better generalization.
- Addition of dropout layer: A new dropout layer with a rate of 0.5 was added after the Dense 128 layer to maintain regularization throughout the network.

These modifications led to significant improvements in the model performance. Model 2, the refined model, achieved an accuracy of 99.39% on the training dataset and 97.43% on the testing dataset, resulting in an overall accuracy of 99.02%. Table II shows the accuracy values of Model 1 and Model 2. The figures highlight that Model 2 is highly effective in recognizing and classifying various yoga postures. This shows a substantial enhancement over the initial model configuration, indicating effective generalization and robustness in pose detection tasks.

TABLE II. ACCURACY SCORE OF EACH YOGA POSE

Yoga pose	Accuracy	
	Model 1	Model 2
Downdog	95.65%	98.10%
Goddess	92.50%	98.21%
Plank	98.24%	99.32%
Tree	98.46%	99.15%
Warrior2	96.22%	99.65%

The comparison between the predicted and actual poses in Model 2 shows a high level of accuracy. Details of this can be seen in Figure 5. This strong agreement between actual and predicted values shows how well the model can validate the correctness of yoga poses. Table III provides a snapshot of the model performance across different poses. This real-time validation not only highlights the model's effectiveness but also demonstrates its adaptability in diverse settings, reinforcing its value as a dependable resource for automated yoga posture detection systems. Such real-world applicability suggests that the model could be effectively integrated into both personal and professional yoga practice settings, offering accurate and consistent feedback to practitioners. Figure 6 presents graphs that depict model accuracy and model loss.

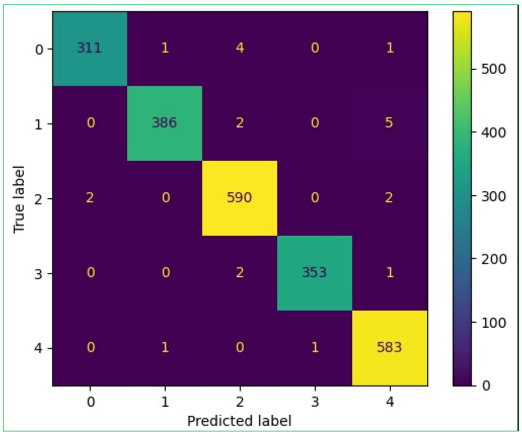


Fig. 5. Confusion matrix of yoga posture recognition.

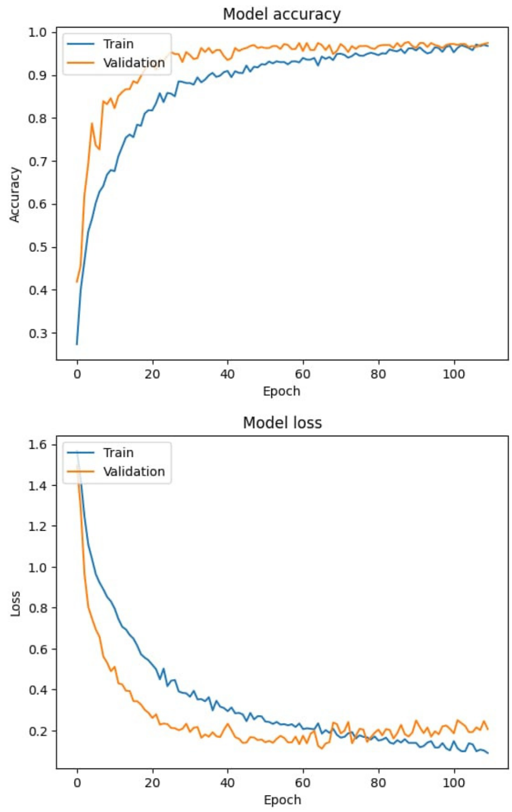
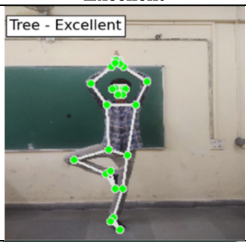
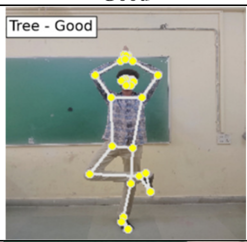
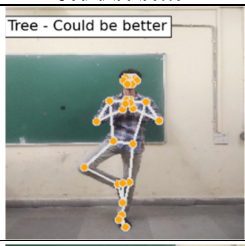
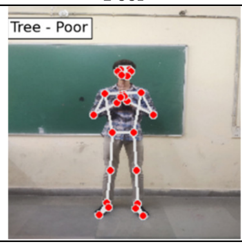
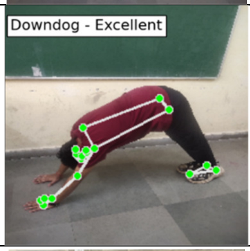
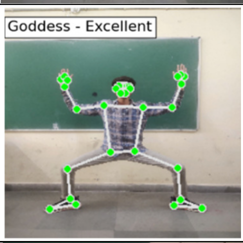
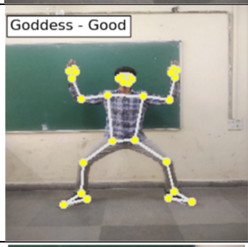
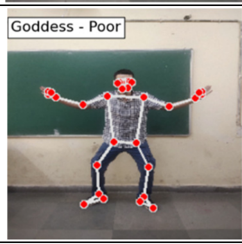
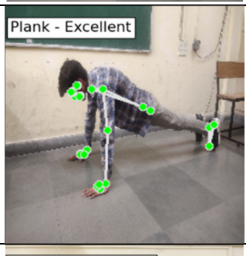
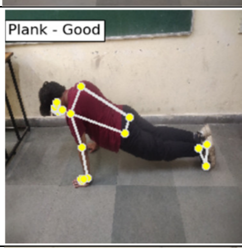
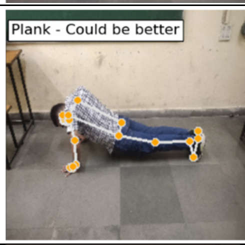
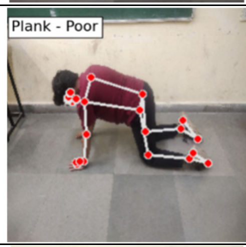
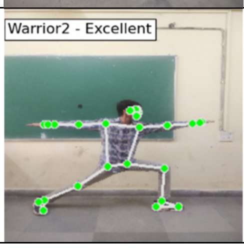


Fig. 6. Model accuracy and loss during training and testing.

TABLE III. EVALUATION OF DETECTED YOGA POSES AND FEEDBACK CATEGORIES

Yoga pose	Excellent	Good	Could be better	Poor
Tree				
Downdog				
Goddess				
Plank				
Warrior2				

IV. CONCLUSION AND FUTURE SCOPE

This study proposed a robust deep-learning approach for yoga pose recognition and personalized feedback in real-time using body landmarks. DeepYoga, the proposed system, achieves high accuracy in classifying the five above-mentioned yoga poses, providing immediate and actionable feedback to the users to improve their posture. This model achieved an average accuracy of 99.02% in detecting and evaluating yoga poses, outperforming the accuracy reported in other studies in this field. The novelty of the proposed system lies in its lightweight design and its ability to provide real-time feedback to practitioners on how closely they match the perfect pose of each exercise. By empowering yoga practitioners to execute postures with greater precision, the DeepYoga system contributes to improving the effectiveness of yoga practice and promoting overall well-being.

Future work could explore several enhancements. First, expanding the dataset with different Yoga poses and diverse body types can make the model more generalized. Additionally, some advanced techniques can be integrated, such as fine-tuning pre-trained models and transfer learning, to further increase the model performance. Furthermore, DeepYoga can be integrated into a web or mobile application to make it more accessible and user-friendly, promoting healthier and more effective yoga practices.

REFERENCES

[1] A. Ross and S. Thomas, "The Health Benefits of Yoga and Exercise: A Review of Comparison Studies," *The Journal of Alternative and Complementary Medicine*, vol. 16, no. 1, pp. 3–12, Jan. 2010, https://doi.org/10.1089/acm.2009.0044.

[2] M. S. Talukder, R. Chiong, Y. Bao, and B. Hayat Malik, "Acceptance and use predictors of fitness wearable technology and intention to recommend," *Industrial Management & Data Systems*, vol. 119, no. 1, pp. 170–188, Jan. 2019, https://doi.org/10.1108/IMDS-01-2018-0009.

[3] Y. Said, M. Barr, and H. E. Ahmed, "Design of a Face Recognition System based on Convolutional Neural Network (CNN)," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5608–5612, Jun. 2020, https://doi.org/10.48084/etasr.3490.

[4] P. Chakraborty and C. Tharini, "Pneumonia and Eye Disease Detection using Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5769–5774, Jun. 2020, https://doi.org/10.48084/etasr.3503.

[5] S. Kale, N. Kulkarni, S. Kumbhkarn, A. Khuspe, and S. Kharde, "Posture Detection and Comparison of Different Physical Exercises Based on Deep Learning Using Media Pipe, Opencv," *International Journal of Scientific Research in Engineering and Management*, vol. 7, no. 4, Apr. 2023.

[6] G. G. Chiddarwar, A. Ranjane, M. Chindhe, R. Deodhar, and P. Gangamwar, "AI-Based Yoga Pose Estimation for Android Application," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 9, pp. 1070–1073, Oct. 2020, https://doi.org/10.38124/IJISRT20SEP704.

[7] I. Chaudhary, N. Thoiba Singh, M. Chaudhary, and K. Yadav, "Real-Time Yoga Pose Detection Using OpenCV and MediaPipe," in *2023 4th International Conference for Emerging Technology (INCET)*, Belgaum, India, May 2023, pp. 1–5, https://doi.org/10.1109/INCET57972.2023.10170485.

[8] H. T. Chen, Y. Z. He, C. C. Hsu, C. L. Chou, S. Y. Lee, and B. S. P. Lin, "Yoga Posture Recognition for Self-training," in *MultiMedia Modeling*, vol. 8325, C. Gurrin, F. Hopfgartner, W. Hurst, H. Johansen, H. Lee, and N. O'Connor, Eds. Cham, Switzerland: Springer International Publishing, 2014, pp. 496–505.

[9] D. Parashar, O. Mishra, K. Sharma, and A. Kukker, "Improved Yoga Pose Detection Using MediaPipe and MoveNet in a Deep Learning Model," *Revue d'Intelligence Artificielle*, vol. 37, no. 5, pp. 1197–1202, Oct. 2023, https://doi.org/10.18280/ria.370511.

[10] V. Bhosale, P. Nandeshwar, A. Bale, and J. Sankhe, "Yoga Pose Detection and Correction using Posenet and KNN," *International Research Journal of Engineering and Technology*, vol. 9, no. 4, pp. 1290–1293, Apr. 2022.

[11] N. Pandit, "Yoga Poses Dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/niharika41298/yoga-poses-dataset.

[12] "mediapipe - PyPI." [Online]. Available: https://github.com/google/mediapipe.

[13] "OpenCV," *Open Computer Vision Library*. https://opencv.org/.