

YOGA POSTURE DETECTION

Keerthana S¹, Mr.Sujayeendra Rao², Dr. Ananth G S³

PG Student, The National Institute of Engineering, Mysuru, Visveswaraya Technological University, Belagavi,
Karnataka, India¹

Faculty, The National Institute of Engineering, Mysuru, Visveswaraya Technological University, Belagavi, Karnataka,
India^{2,3}

Abstract: Yoga posture correction and recognition systems can help learners practice safely and consistently by providing instant feedback without requiring continuous supervision from an instructor. This project proposes a deep learning-based Yoga Posture Detection system that identifies yoga poses from images and real-time webcam video. The system uses a convolutional neural network (CNN) based classifier trained on a labeled dataset of yoga postures, where each class corresponds to a specific asana. For real-time operation, YOLOv8 is used to detect the person in each frame, the detected region is cropped, and the posture is classified using the trained model. The classification model is trained using transfer learning (MobileNetV2 backbone) to improve accuracy with a limited dataset and reduce training time. The final system is deployed as a web application using Flask with a user-friendly interface built using HTML, CSS, and JavaScript, allowing users to upload images for posture prediction and view top confidence results. Experimental results show that the proposed model achieves around 70% validation accuracy over 41 yoga classes, and performance is analyzed using a confusion matrix, classification report, and Grad-CAM visual explanations. The solution demonstrates an end-to-end pipeline for yoga pose classification and real-time detection, and can be extended further for posture correction and fitness guidance.

Keywords: deep learning, yoga pose detection, Convolutional Neural Network (CNN), Real-Time Detection.

I. INTRODUCTION

In recent years, the integration of artificial intelligence with healthcare and fitness applications has gained significant attention due to its potential to improve accessibility, safety, and personalized guidance. Practicing yoga has always provided a method for people to enhance their physical, mental and/or emotional well-being. Practicing yoga correctly could provide practitioners the greatest benefit no matter what condition they were currently experiencing. Although there are advantages to correctly teaching yogic positions to practitioners as they learn and practice them, there are also many obstacles to successfully learning these positions and then correctly using the postures within their own bodies.

In the past few years, significant advances have been made in the development of Computer Vision, Deep Learning, and Convolutional Neural Networks to assist in resolving these difficulties. Technology can create an automated system that can identify and interpret how a person moves by processing images and video data using Convolutional Neural Networks. With the advent of these systems, it is now possible to achieve a more effective solution to Yoga Pose Detection by combining Human Detection and Classification models using Deep Learning Algorithms. An automated Yoga Pose Detection System allows users to monitor and determine their own Yoga Poses and track and identify the Yoga Poses of all other practitioners using a real-time Object Detection System.

This project is an attempt to aid people looking for immediate information about poses from the practice of Yoga through an AI Model/Web Application. The intention of this project is to show how recent advancements in technology (specifically Artificial Intelligence) can provide a way to self-evaluate and self-assess your Yoga practice safely and effectively. Therefore, the study's goal is not to replace traditional methods of learning, but rather bridge the gap between these two styles of learning.

II. SCOPE OF THE LITERATURE SURVEY

The literature on yoga posture detection demonstrates significant progress in the application of deep learning and computer vision for automated fitness and healthcare systems. Early approaches mainly relied on traditional machine learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and handcrafted feature extraction methods, which showed limitations in handling variations in lighting, body orientation, clothing, and complex backgrounds. Recent studies have shifted towards deep learning-based approaches, especially Convolutional Neural Networks (CNNs), transfer learning, and pose estimation frameworks, which provide better accuracy and robustness in real-world environments. Several researchers have explored CNN-based yoga pose classification using transfer learning

models such as MobileNetV2 due to their lightweight architecture and efficient feature extraction capabilities. Survey studies by Rajendran and Sethuraman highlighted three major research directions in yoga posture recognition: image classification methods, skeleton-based pose recognition, and hybrid feedback systems. Their work also identified common challenges such as dataset imbalance, limited pose diversity, and difficulties in real-time deployment. Research by Bansal et al. focused on real-time yoga pose recognition using landmark-driven CNN models, emphasizing user feedback and live webcam-based prediction. Similarly, transformer-based models such as ViTPose proposed by Xu et al. demonstrated improved human pose estimation by capturing global body relationships and producing accurate keypoints under challenging conditions. Jiang et al. introduced RTMPose, which achieved high real-time performance with low latency, making it suitable for live yoga coaching systems and multi-person detection scenarios. Object detection frameworks have also played an important role in improving system efficiency. Studies on YOLOv8 architecture showed that real-time person detection can significantly improve posture classification by accurately cropping the human region before classification. Additionally, explainable AI techniques such as Grad-CAM have been incorporated in recent systems to generate heatmaps that highlight important body regions influencing model predictions, thereby improving model transparency and user trust. Overall, the literature indicates that combining YOLO-based person detection, CNN-based transfer learning models, pose estimation techniques, and explainable AI methods creates an effective framework for accurate and scalable yoga posture detection systems. These advancements support the development of real-time, user-friendly applications capable of assisting users in posture recognition, fitness monitoring, and personalized yoga guidance.

III. PROPOSED WORK

The proposed system aims to develop an intelligent Yoga Posture Detection System using deep learning and computer vision techniques to accurately recognize yoga poses from both static images and real-time webcam video. The system is designed to assist users in practicing yoga independently by providing automated posture recognition with confidence scores through a user-friendly web application. The proposed work utilizes a Convolutional Neural Network (CNN) based classification model with Transfer Learning using the MobileNetV2 architecture. MobileNetV2 is selected because of its lightweight structure, faster computation, and ability to achieve high accuracy even with limited training data. The model is trained on a dataset containing 41 yoga posture classes with approximately 2325 images. Image preprocessing techniques such as resizing, normalization, RGB conversion, and cropping are applied to improve model consistency and prediction performance. For real-time functionality, the system integrates YOLOv8 object detection to detect the person from webcam frames. The detected person region is cropped and passed to the CNN classification model for yoga posture prediction. This two-stage pipeline improves classification accuracy by focusing only on the human body and removing unnecessary background information. The system also supports top-k predictions with confidence scores for better interpretation of results. To improve transparency and interpretability, the proposed system incorporates Grad-CAM visualization, which generates heatmaps indicating the important body regions influencing the model's predictions. This helps users understand how the model identifies yoga postures and increases trust in the prediction process. The entire system is deployed as a Flask-based web application using HTML, CSS, and JavaScript for the frontend interface. Users can upload images, provide image URLs, or use real-time webcam detection to identify yoga poses instantly. The modular architecture of the system also allows future enhancements such as pose correction, landmark-based feedback, personalized yoga recommendations, and difficulty-level estimation. Overall, the proposed work focuses on developing a scalable, efficient, and real-time yoga posture recognition system that combines deep learning, transfer learning, object detection, and explainable AI techniques for practical fitness and healthcare applications.

IV. METHODOLOGY

The proposed Yoga Posture Detection system is developed using a deep learning and computer vision-based methodology that combines image preprocessing, transfer learning, object detection, real-time prediction, and web deployment. The methodology is designed to provide accurate yoga posture recognition from both static images and live webcam streams.

- [1] **Dataset Collection and Preparation:** The dataset used in this project consists of approximately 2325 images categorized into 41 different yoga posture classes. Each class is stored in a separate folder representing a specific yoga asana. The dataset contains images captured under different lighting conditions, backgrounds, body orientations, and clothing styles to improve model generalization. The dataset is divided into training and validation sets using an 80:20 ratio.
- [2] **Data Preprocessing:** Preprocessing is performed to convert raw images into a standardized format suitable for deep learning models. All images are resized to 224×224 pixels to match the input requirements of the MobileNetV2 architecture. Pixel values are normalized to the range $[0,1]$ using rescaling techniques. RGB conversion, cropping, and aspect ratio preservation are also applied to improve prediction consistency. During real-time detection, YOLOv8-generated bounding boxes are used to crop the detected person region before classification.

- [3] **Model Selection Using Transfer Learning:**The proposed system uses a Convolutional Neural Network (CNN) with Transfer Learning based on the MobileNetV2 architecture. MobileNetV2 is selected because it is lightweight, computationally efficient, and suitable for real-time applications. The pretrained ImageNet weights are utilized as the feature extraction backbone. The top classification layers are removed and replaced with:
- Global Average Pooling Layer
 - Dense Layer with 256 neurons and ReLU activation
 - Output Dense Layer with 41 neurons and Softmax activation
- The Softmax layer produces probability scores for all yoga posture classes.
- [4] **Model Training and Validation:**The model is trained using the Adam optimizer and categorical cross-entropy loss function, which is suitable for multi-class classification problems. Training is performed with a batch size of 32 for 20 epochs. During training, validation accuracy and loss are monitored to evaluate model performance and generalization capability. The trained model is saved in .h5 format for deployment and future inference.
- [5] **Real-Time Person Detection Using YOLOv8:**For webcam-based posture detection, YOLOv8 is integrated as the object detection module. YOLOv8 detects the “person” object from each webcam frame and generates bounding boxes around the detected human body. The detected region is cropped and passed to the trained CNN classifier for posture prediction. This approach improves accuracy by removing unnecessary background information and focusing only on the human posture.
- [6] **Pose Classification and Prediction:**The cropped image is preprocessed and passed to the trained MobileNetV2 classification model. The model predicts the yoga posture label along with confidence scores. The system also supports top-k predictions, enabling users to view multiple probable posture classes ranked by prediction probability.
- [7] **Model Evaluation:**The performance of the proposed system is evaluated using multiple metrics such as validation accuracy, precision, recall, F1-score, and confusion matrix analysis. The system achieved approximately 80.60% validation accuracy across 41 yoga posture classes. Grad-CAM visualization is also used to generate heatmaps that highlight important body regions influencing the prediction, thereby improving explainability and interpretability.
- [8] **Web Application Deployment:**The final system is deployed using the Flask framework with HTML, CSS, and JavaScript for frontend development. The web application allows users to upload yoga images, provide image URLs, or use live webcam detection for posture prediction. The backend processes the image, performs classification, and returns posture labels with confidence scores in real time.
- [9] **Overall Workflow:**The complete workflow of the system includes:
1. Image/Webcam Input
 2. Image Preprocessing
 3. YOLOv8 Person Detection (Real-Time Mode)
 4. Region Cropping and Normalization
 5. MobileNetV2-Based CNN Classification
 6. Prediction Generation with Confidence Scores
 7. Result Display through Web Interface

This methodology provides an efficient, scalable, and real-time solution for yoga posture recognition and forms a strong foundation for future extensions such as posture correction, personalized fitness guidance, and pose difficulty estimation.

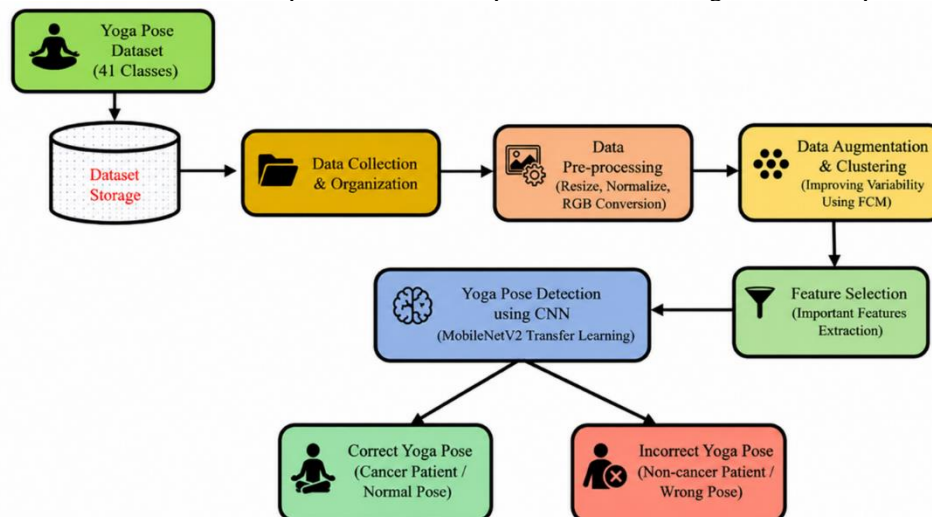


Fig 1. Workflow of the project

VI. RESULT ANALYSIS

The proposed Yoga Posture Detection system was evaluated using multiple performance metrics to measure the effectiveness of posture classification and real-time detection. The system was trained using the MobileNetV2 transfer learning model with 41 yoga posture classes and tested on unseen validation data. The overall validation accuracy achieved by the system was approximately 80.60%, indicating that the model was able to correctly classify most yoga poses with good generalization capability. During training, the model showed a steady improvement in accuracy while the training loss gradually decreased, demonstrating effective learning of yoga posture features. The training accuracy reached nearly 99%, while the validation accuracy stabilized around 80%, suggesting slight overfitting due to similarities between yoga poses and class imbalance within the dataset. However, the use of transfer learning, normalization, and preprocessing techniques helped improve overall model performance. The classification report provided detailed class-wise analysis using Precision, Recall, and F1-Score metrics. Yoga poses with distinct body structures and movements achieved high precision and recall values because the model could easily differentiate them from other poses. In contrast, visually similar poses showed comparatively lower performance due to overlapping body alignments and insufficient training samples. This analysis highlights the importance of balanced datasets and pose diversity in improving prediction accuracy. The Confusion Matrix analysis further demonstrated how certain yoga postures were misclassified into similar classes. Poses involving comparable hand, leg, or sitting positions showed higher confusion rates. The confusion matrix helped identify weak classes and provided insight into areas where additional training data and augmentation techniques could improve the system. For explainability, the system utilized Grad-CAM visualization, which generated heatmaps highlighting the body regions that influenced model predictions. The heatmaps showed that the CNN model primarily focused on arms, legs, torso orientation, and body posture while identifying yoga poses. This improved the interpretability and transparency of the deep learning model. The real-time webcam detection module integrated with YOLOv8 successfully detected human subjects from live video streams and accurately cropped the person region before classification. The system demonstrated smooth real-time performance with acceptable latency and stable frame processing. The live detection system was able to overlay predicted posture labels and confidence scores on the video feed effectively. The web-based Flask application also functioned successfully by allowing users to upload images, provide image URLs, and obtain posture predictions instantly. API testing, validation checks, image upload handling, and webcam integration tests were successfully passed, confirming the reliability and robustness of the developed system. Overall, the experimental results demonstrate that the proposed Yoga Posture Detection system effectively combines deep learning, transfer learning, object detection, and real-time computer vision techniques to achieve accurate yoga posture recognition. The system provides a scalable and user-friendly solution that can be further enhanced with pose correction modules, larger datasets, and personalized fitness guidance features in future work.

Image Prediction

Drop a photo or choose an image — the system will detect the yoga pose and show detailed benefits.



The screenshot displays a web interface for image prediction. On the left, there is a placeholder for an image, which has been replaced by a photograph of a person performing a Big Toe Pose. Above the image are three buttons: 'Choose Image' (green), 'Predict' (blue), and 'Clear' (grey). To the right of the image, a list of predicted yoga poses is shown with corresponding progress bars and confidence percentages:

Yoga Pose	Confidence
padangusthasana	92%
uttanasana	6%
halasana	1%
gomukhasana	1%
utthita parsvakonasana	0%

Below the list, the system identifies the pose as **Big Toe Pose** and provides a description: "Fold forward and hold the big toes with fingers; hinge from the hips and keep a long spine. Bend knees slightly if hamstrings are tight." It also lists the **Benefits**: "Deep hamstring stretch that calms the nervous system and improves flexibility in the calves and lower back."

CONCLUSION

This project successfully developed and demonstrated an end-to-end Yoga Posture Detection system using deep learning and computer vision techniques. The main objective was to recognize yoga postures from images and real-time video and provide an accessible interface for users through a web application. To achieve this, a labeled dataset of yoga images was organized into class-wise folders and analyzed for sample distribution. The dataset contained 41 yoga pose classes with approximately 2325 images, and an 80:20 split was applied for model training and validation. MobileNetV2 is a deep network architecture which was leveraged in constructing a transfer-learning based CNN Model for Posture Classification. The CNN Model was developed by conducting 20 epochs, resulting in an approximate accuracy of 80.60% on the validation dataset. The training process for the CNN Model was characterised by multiple iterations, generating diverse outputs by generating multiple sequences through applying data augmentation and dropout techniques. Hence, the output of the CNN Model was extremely varying or possibly random based on the yoga pose being classified. However, the method developed by using the MobileNetV2 architecture provided a way to effectively distinguish between the different poses. The visualisation (Feature Visualization) and Grad-CAM analysis of the model also demonstrated which parts of an image had an impact on generating a particular classification and created an opportunity to define the classification method more accurately. Through these processes, it became apparent that the training of different models would be necessary for different image classifying tasks. The application of the YOLOv8 model allowed us to implement user detection in video from a webcam and to identify areas in the user's field of view that are important for the recognition of yoga poses. The YOLOv8 model also allowed us to eliminate background noise when doing so. The YOLOv8 application was implemented with the Flask framework. For user interaction with this model, we developed web based applications using HTML, CSS, and JavaScript. The implementation of this project has clearly demonstrated that a successful and effective model for recognizing all forms of yoga poses and developing training/testing methodologies for these poses can be developed with the use of the YOLOv8 model.

VII. ACKNOWLEDGMENT

I extend my deepest gratitude to Mr. Sujayeendra Rao, project supervisor, for her expert guidance, constructive feedback, and steadfast support throughout this research. Sincere appreciation is due to the faculty and staff of the National Institute of Engineering, Mysuru, for facilitating critical resources and institutional infrastructure. Heartfelt thanks go to peers and classmates for their collaborative input and motivation, and to my parents for their unwavering emotional and moral encouragement. Finally, acknowledgment is given to all who contributed directly or indirectly to this project's successful completion.

REFERENCES

- [1]. Bansal, R., Sharma, R., Jain, P., Arora, R., Pal, S. and Vishal (2024) 'Deep Yoga: Enhancing Practice with a Real-Time Yoga Pose Recognition System', *Engineering, Technology & Applied Science Research*, 14(6), pp. 17704–17710. ETASR
- [2]. Borthakur, D., Paul, A., Kapil, D. and Saikia, M.J. (2023) 'Yoga Pose Estimation Using Angle-Based Feature Extraction', *Healthcare*, 11(24), 3133. MDPI
- [3]. Jiang, T., Lu, P., Zhang, L., Ma, N., Han, R., Lyu, C., Li, Y. and Chen, K. 'RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose', *arXiv preprint arXiv:2303.07399*. arXiv
- [4]. Kishore, P.V.V., et al. (2022) 'Estimation of Yoga Postures Using Machine Learning Techniques', *International Journal of Yoga*. PubMed
- [5]. Parashar, A., et al. (2023) 'Improved Yoga Pose Detection Using MediaPipe and MoveNet in a Deep Learning Model', *Revue d'Intelligence Artificielle*, 37(5)