Human Pose Estimation using Machine Learning

Mrinmayee Deshpande[1], Dipali Gangarde[1], Nishchay Bhardwaj[1], Ashish Kumar Yadav[1], Nilesh P. Sable[1], Anuradha Yenkikar[1]

Department of CSE - Artificial Intelligence, Vishwakarma Institute of Information Technology, Pune, India
mrinmayee.22210087@viit.ac.in

Department of CSE - Artificial Intelligence, Vishwakarma Institute of Information Technology, Pune, India
dipali.22210638@viit.ac.in

Department of CSE - Artificial Intelligence, Vishwakarma Institute of Information Technology, Pune, India
nishchay.22211583@viit.ac.in

Department of CSE - Artificial Intelligence, Vishwakarma Institute of Information Technology, Pune, India
ashishkumar.22210133@viit.ac.in

Department of CSE - Artificial Intelligence, Vishwakarma Institute of Information Technology, Pune, India
drsablenilesh@gmail.com

Department of CSE - Artificial Intelligence, Vishwakarma Institute of Information Technology, Pune, India
anu.jamkhande@gmail.com

Abstract: The human pose estimation proves useful for various tasks such as healthcare, sports analysis and human computer interaction. The yoga practitioners have gained wide momentum in today's era due to increasing health awareness. With the rising popularity of yoga and the increasing demand for technology-assisted learning platforms, there is an increasing need for accurate and effective methods for recognizing yoga poses. Yoga pose estimation and classification plays a crucial role in automated yoga training systems, enhancing the accessibility and effectiveness of yoga practice. Accurate recognition of yoga poses is essential for safe and effective yoga practice. Incorrect posture can result in serious injury to the body, emphasizing the critical need for precise pose detection and classification. This concern motivated our research to explore automated systems for yoga pose recognition. The paper aims to facilitate accurate identification of yoga poses, thereby enhancing accessibility to yoga instruction and minimizing the risk of injury associated with improper form. This research paper focuses on the comparison of two models namely, Ultralytics and MoveNet for detecting the keypoints in yoga poses. The keypoints are subsequently utilized for classification into five different yoga poses: downdog, goddess, plank, tree, and warrior2. Evaluation shows MoveNet achieving a superior accuracy of 93% compared to Ultralytics 88%. Precision, recall, and F1 scores are analyzed through confusion matrices for a performance analysis. This study advances automated yoga pose recognition, providing insights into the capabilities and limitations of current deep learning approaches.

Keywords: YOLO-v8, MoveNet, Ultralytics, Yoga, HealthCare.

1. Introduction

Yoga integrates the body, breath, and meditation to improve mental, emotional, and spiritual well-being. Yoga is widely practiced worldwide due to its numerous health advantages, such as lowering stress, boosting flexibility, and enhancing muscular tone. We have seen a sharp increase in the use of technology for yoga practice in recent years. Due to the convergence of technology and wellness, there has been a noticeable surge in the popularity of yoga worldwide in recent years. Because of the increased focus on fitness and health, as well as the accessibility of technology-assisted learning platforms, more individuals are turning to yoga as a means of enhancing their physical and mental well-being. However, as yoga gains popularity, there will be a larger demand for accurate and practical methods for identifying and rating yoga poses. Yoga is an ancient Indian form of exercise that involves a vast range of poses, or "asanas," each designed to target a certain muscle group and improve flexibility and overall body awareness. Even while yoga has many benefits, there is still a risk of injury due to improper alignment or technique, particularly for beginners or those without access to trained instructors. Since correct pose detection and categorization is crucial, our goal is to research automated systems for yoga position recognition. By leveraging developments in computer vision and deep learning to enable accurate recognition of yoga postures, we hope to increase accessibility to yoga education and lower the risk of harm associated with bad posture. Our research focuses on the extracting keypoints of the body and then identifying the yoga pose accurately. The journey begins with collection of dataset. We have used the dataset having the yoga images for achieving the goal of our paper. This dataset is publicly available on Kaggle. The keypoints such as NOSE, LEFT_EYE, RIGHT_EYE, LEFT_EAR, RIGHT_EAR, LEFT_SHOULDER, RIGHT_SHOULDER, LEFT_ELBOW, RIGHT_ELBOW, LEFT_WRIST, RIGHT_WRIST, LEFT_HIP, RIGHT_HIP, LEFT_KNEE, RIGHT_KNEE, LEFT_ANKLE, RIGHT_ANKLE were extracted for identifying pose. The algorithm used here is YOLO V-8. The dataset consists of two folders, namely, Train and Test. Five pose labels such as down-dog, tree, goddess, plank, warrior2 were used for the classification. The model was assessed using a variety of parameters including recall, accuracy, and f1-score. Our primary objective is to create a model capable of accurately identifying and categorizing yoga positions in real time. We preprocessed the dataset to normalize the pose photos and then identify the important spots in order to achieve this goal. We used a training dataset to train Yolo v-8 and MoveNet, and we assessed their performance at the checkout set using a variety of evaluation metrics in addition to F1-rating, precision, and recall. This work aims to develop automated systems for the recognition of yoga poses through illuminating the efficacy of various models and their ability to accurately detect yoga postures. Ultimately, our study attempts to provide people with the knowledge and abilities needed to practice yoga safely and efficiently, encouraging a more approachable and healthful approach to wellness.

2. Literature Survey

The importance of practicing yoga is increasing day by day. Nearly 10 papers were surveyed by us. The information gained from these papers is given as follows: Images made up the dataset used in reference [1]. The dataset used in the study consisted of RGB images of five different yoga poses like downdog, tree, plank, warrior2, and goddess—that were manually labelled using the Makesense AI picture annotation tool. The primary results showed the employed model had 93.9% PCK in the goddess position. The study developed an algorithm to evaluate the precision of body joint recognition in yoga poses. The algorithm reached a maximum PDJ of 90% to 100% for most body joints and a high PCK of 93.9% for the goddess posture. The algorithm used was Mediapipe Blazepose Model.

An open-source dataset comprising video footage of fifteen volunteers executing six distinct yoga poses was used in the study in [2]. Fuzzy logic- based modelling is recommended as the feature extraction method for human action recognition. The article describes a modelling technique for human action detection based on fuzzy logic, evaluates the system's performance using benchmark datasets, and extracts features using fuzzy membership functions. The Blazepose model is used by Mediapipe to identify 33 body. The study employs the Blazepose model to identify 33 body keypoints and the Mediapipe and Angle heuristic technique to classify yoga asanas. [3] On the Yoga-82 dataset used in [4], the proposed deep convolutional neural network model, YPose, demonstrated state-of-the-art performance. The model outperforms the previous state-of-the-art with an accuracy of 93.28%, which was 79.35%, by about 13.9%. According to [5], the technique is suitable for a range of applications and is anticipated to benefit the yoga sector. It recognises yoga poses in real time using OpenPose for 3D joint mapping. It is envisaged that the proposed system will benefit the yoga industry by providing an accurate, affordable, and efficient method of position identification. For yoga position categorization, the suggested wavelet-based CNN model in [6] performs better than other comparison models, such as standard CNN models and transfer learning techniques like ResNet18, ResNet50, and GoogleNet. When deep learning is used to classify yoga poses, using wavelet subbands yields better accuracy results than training on the original images or individual subbands. The suggested approach exhibits encouraging accuracy percentages and shows how wavelet decomposition may be used to increase the precision of yoga pose classification. The application of motion analysis in an exergame engine for feedback in collegiate yoga practice, the use of deep learning for human posture recognition in yoga practice, and the suggestion of an Internet of Things-based system for privacy-preserving yoga pose detection are among the primary discoveries in [7]. The work in [8] focuses on the detection and correction of yoga poses using computer vision technologies for human posture assessment. The process is feeding the input image through a CNN classifier that has been trained to detect faces, recognise human body postures, and search for joints and limbs that have been previously taught in order to provide the user with markers that indicate different body parts. The study [9] suggests XGBoost, a real time machine learning framework that outperforms earlier models in terms of accuracy, latency, and size for accurate yoga stance identification. In the work presented in [10], the accuracy of different regression and classification algorithms was compared utilising a person's stance skeleton in photos as a means of recognising and classifying human activities. The authors of the study created the dataset, which is split into two sections for training and validation. The multivariate logistic regression method is used to complete the activity classification task. They have employed the DenseNet architecture in paper [11] to leverage the hierarchical labelling for better posture recognition. In order to address the lack of difficulty and diversity in current pose datasets, especially those created primarily for large-scale yoga pose recognition, it discusses the idea of fine-grained hierarchical pose categorization and offers a new dataset named Yoga-82 with 82 yoga posture classes.

The dataset is distinctive in that it emphasises the variety and complexity of human positions in yoga-related activities. In the study [12], a self-assisted system for identifying and categorising yoga poses was created using computer vision. In order to categorise stances into yoga asanas, the system evaluates real-time video data. It then shows the identified asana together with a confidence score. The goal of the work in [13] was to apply deep learning-based techniques for precise estimation of yoga positions. Among the four deep learning architectures that were employed, MediaPipe had the highest estimation accuracy.

With a high average confidence score of 92%, the study in [14] presented an interactive method for identifying yoga positions using Kinect technology. In paper [15], Convolutional and Recurrent Neural Networks are combined in deep hybrid architecture to estimate a person's pose. The main discovery is the creation of a hybrid architecture combining CNNs and RNNs to provide an endto-end method for estimating human position. Various algorithms which were used in the literature surveyed include – Mediapipe Blazepose Model, Media pipe pose estimation library, support vector machine classifier, random forest classifier, k-nearest neighbours classifier, logistic regression, naïve-Bayes classifier, CNN, OpenPose, MATLAB R2021a, Computer vision technology, XGBoost Classifier, etc.

3. Proposed Methodology

The yoga pose estimation has gained a wide momentum nowadays as the health concern in people is increasing. The relevance of yoga pose detection is that the pose should be detected accurately as the wrong pose can cause serious damage to the body. The paper focuses on the comparison of two models which are accessed on the criteria: accuracy, recall, precision, and F1-score. Numerous pretrained models are available for detecting the keypoints like BlazePose, OpenPose, ResNet, MoveNet, MediaPipe, pose estimation model of ultralytics. Of them, we have chosen the models movenet and pose estimation model of ultralytics.

3.1 Dataset

The dataset which we used is obtained from kaggle. The dataset contained two folders named ‗Train‘ and ‗Test‘. Train folder has 1081 images and test folder has 470 images. Overall, 5 poses namely, goddess, downdog, plank, tree and warrior2 form the classes of the dataset [16] as it is evident from figure 1. Distribution of 5 poses across the dataset is shown in figure 2.

(a) Warrior2 (b) Plank (c)Downdog

(d) Tree (e)Goddess

Fig.1. Sample Images from dataset

Furthermore, we have also implemented real time for estimation. We are developing our real-time implementation in Python with the TensorFlow framework. We make use of TensorFlow's high- level APIs and effective model serving features to provide a smooth integration of the MoveNet model into our application. OpenCV isrequired for image processing and visualisation, among other things.

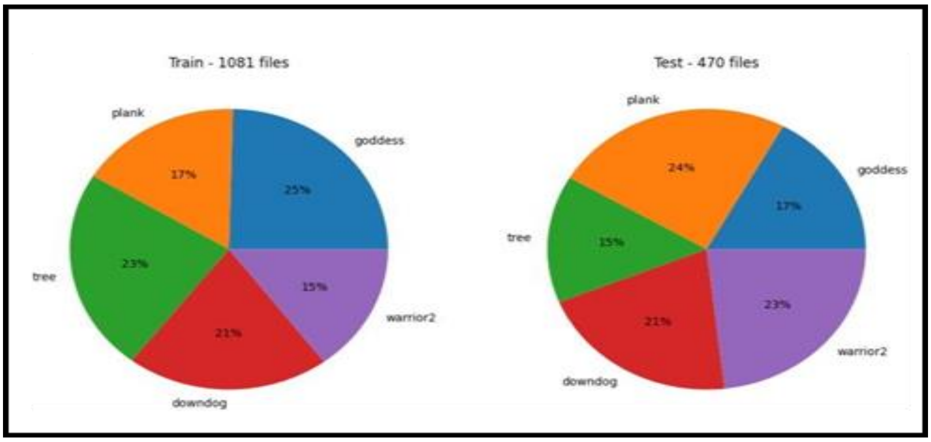


Fig.2. Distribution of all classes of pose

4. Model Architecture

4.1 Approach 1 YOLO-V8 Ultralytics

In this approach, we utilize the YOLOv8 architecture for detecting keypoints corresponding to the yoga poses in the images. YOLOv8 (You Only Look Once version 8) is a state of-the-art object detection model known for its efficiency and accuracy in detecting objects and keypoints in images. The preprocessing steps includes resizing the images, conversion of images from BGR to RGB. The pretrained model for pose estimation, ultralytics was used to detect the keypoints. In total, 17 keypoints were detected. After the extraction of all the keypoints, the keypoints of each body part were stored in a CSV file. CSV file included image name, class label and remaining column for each body keypoints. A custom neural network classifier is designed for classification of yoga poses based on the detected keypoints. The label encoder was used to label the classes. This classifier is responsible for mapping the extracted keypoint information to the corresponding yoga pose labels. Multiple convolutional layers are the first layer in the classifier‘s design and these are followed fully linked layers. Convolutional layers are in charge of feature extraction, capturing relevant patterns and relationships within the keypoint data. For pose categorization, these attributes are then flattened and run through fully connected layers. The neural network's hidden layers employ the Rectified Linear Unit (ReLU) activation function. ReLU adds non-linearity to the data, enabling the model to discover intricate linkages therein. For multi-class classification, the output layer applies the Softmax activation function. It converts the raw output scores into probabilities, representing the likelihood of each yoga pose class. For gradient descent optimization, Adam optimizer is employed. The Adam optimizer is renowned for its potency in deep neural network training. Based on the loss function's gradients in relation to the model's parameters, it dynamically modifies the learning rate. We implemented a learning rate of 0.01, a batch size of 12, and trained the model for 40 epochs to ensure balanced optimization and efficient convergence. Graphical Processing Unit(GPU) is required to run the model. This adaptive learning rate helps in faster convergence and improved performance during training.

The classification report as obtained for the ultralytics model is as shown in fig(3) and it's graphical representation is in figure(4) –

	precision	recall	f1-score	support
downdog	0.95	0.93	0.94	40
goddess	0.97	0.77	0.86	39
plank	0.82	0.98	0.89	51
tree	0.82	0.88	0.85	32
warrior2	0.87	0.82	0.85	57
accuracy			0.88	219
macro avg	0.89	0.87	0.88	219
weighted avg	0.88	0.88	0.88	219

Fig. 3. Classification report for ultralytics and neural network classifier

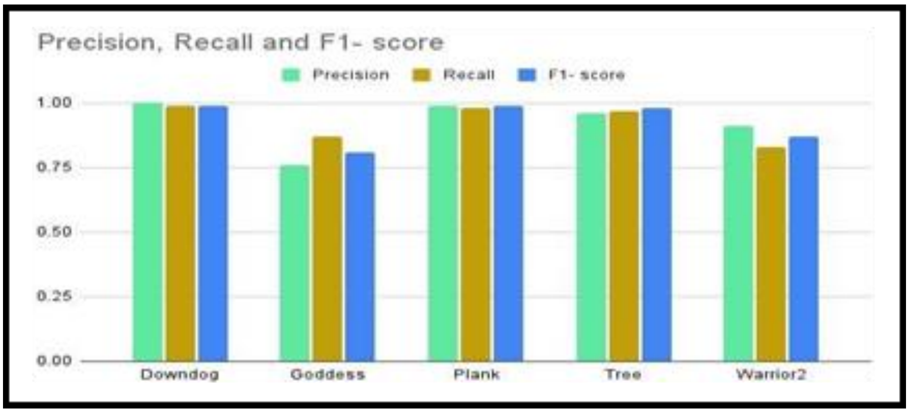


Fig. 4. Graphical View of Classification report for ultralytics and neural network classifier

4.2 Approach 2 - MoveNet

MoveNet is a compact deep learning network specifically created for efficient and precise identification of keypoints, such as human posture estimate. There are two types of MoveNet models namely, MoveNet- lightning and Movenet- Thunder. In our study, we utilize MoveNet lightning for detecting 17 keypoints corresponding to human body joints in yoga poses. These keypoints include joints such as shoulders, elbows, wrists, hips, knees, and ankles, which are essential for accurately representing the body posture in different yoga poses. The preprocessing steps include resizing images into shape 192 pixels, conversion of images from BGR to RGB. After detecting the keypoints using

MoveNet, we employ a separate neural network classifier to classify the yoga poses based on the detected keypoints. This classifier is responsible for mapping the extracted keypoint information to the corresponding yoga pose labels. The architecture of the classifier follows a similar structure to the YOLOv8-based classifier, which includes convolutional layers followed by fully connected layers. However, specific adjustments may be made to accommodate the input size and features extracted by MoveNet. Convolutional layers are in charge of feature extraction, capturing relevant patterns and relationships within the keypoint data. After being flattened, these features are routed via fully connected layers in order to classify poses. Similar to the YOLOv8-based approach, appropriate activation functions, loss functions, and optimization algorithms are utilized for training the classifier. ReLU (Rectified Linear Unit) activation function is frequently employed in the hidden layers of the neural network for introducing non-linearity and facilitating convergence during training. For multi-class classification, Softmax activation function is employed in the output layer providing normalized probabilities for each yoga pose class. For multi-class classification tasks, the loss function of choice is frequently cross-entropy loss.

The dissimilarity between the expected probability distribution and the ground truth labels is quantified by this measurement. Adam optimizer, is commonly used for gradient descent optimization. for the MoveNet model, we applied a batch size of 32, and trained for 40 epochs, achieving smooth training dynamics and effective model refinement. These settings ensure that the classifier effectively learns to identify and classify yoga poses based on the keypoints detected by MoveNet, ultimately leading to accurate and reliable pose recognition results.
The classification report as obtained for the MoveNet model is as shown in figure 5 and it’s graphical representation is in figure 6 –

	precision	recall	f1-score	support
0	1.00	0.99	0.99	91
1	0.76	0.87	0.81	77
2	0.99	0.98	0.99	108
3	0.96	0.97	0.96	66
4	0.91	0.83	0.87	104
accuracy			0.93	446
macro avg	0.92	0.93	0.92	446
weighted avg	0.93	0.93	0.93	446

Fig. 5. Classification report for MoveNet and neural network classifier

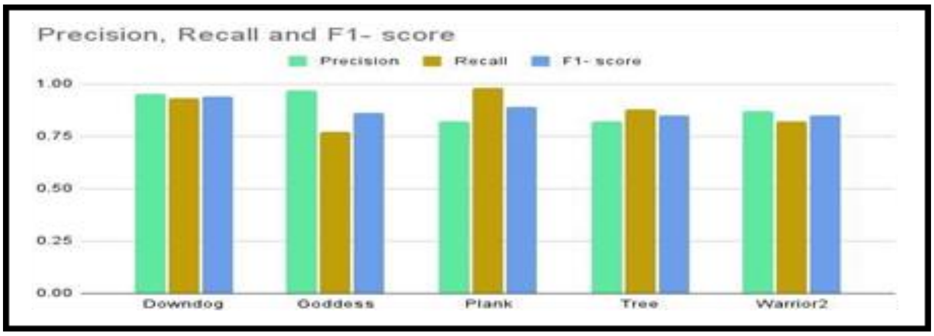


Fig. 6. Graphical View of Classification report for MoveNet and neural network classifier

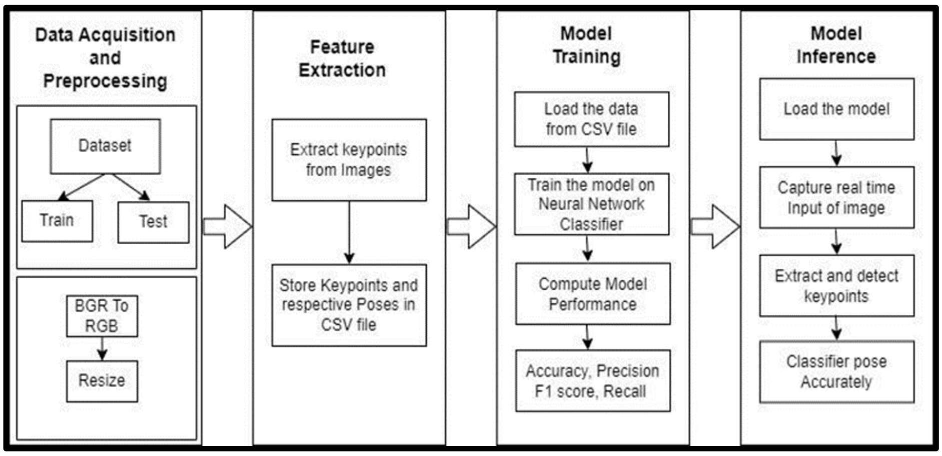


Fig. 7. Architecture of Model

5. Results

The accuracy of the two models' performance—MoveNet and YOLO-v8—was assessed in relation to their ability to identify yoga poses. **Figs. 8 and 9** show the training, validation loss for both the models.

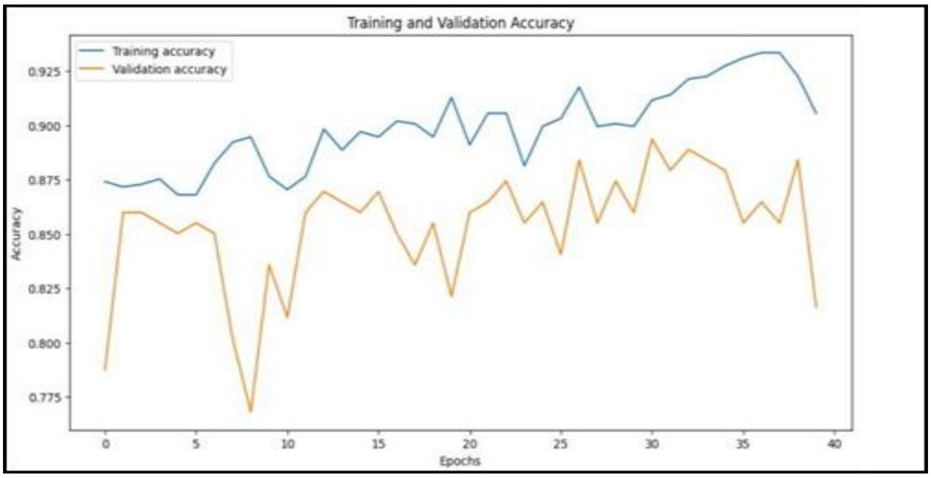


Fig. 8. MoveNet model's training and validation accuracy

Fig. 9. MoveNet model's training and validation loss

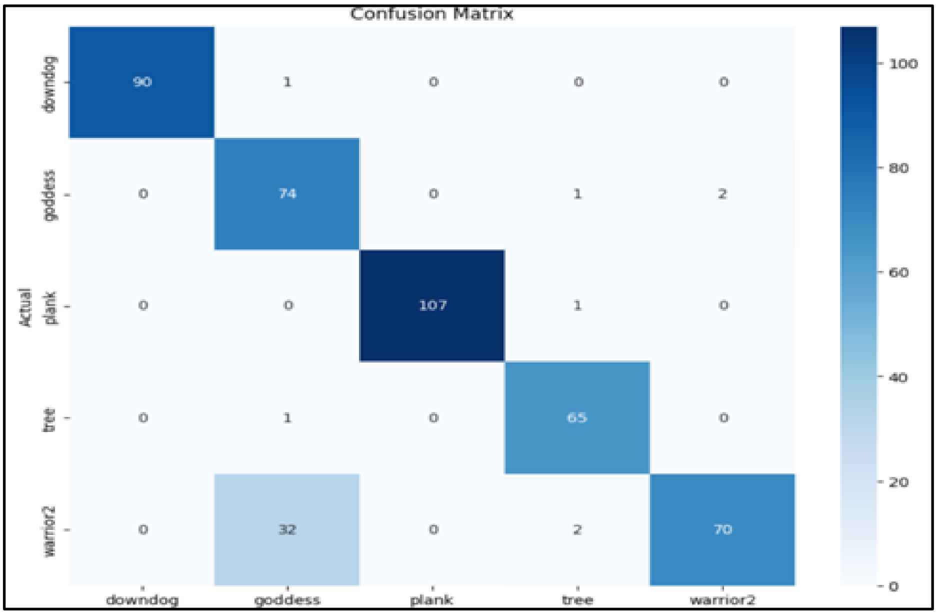


Fig. 10. Confusion Matrix (MoveNet Model)

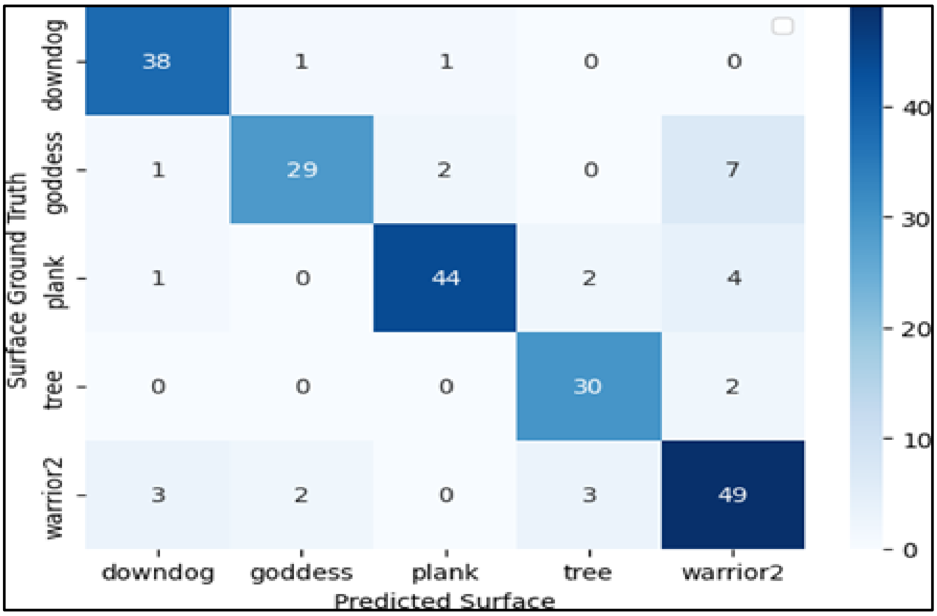


Fig.11. Confusion Matrix (Yolov8 ultralytics Model)

Figs. 10 and 11 show the confusion matrices for both the models. As shown in Fig. 10, the confusion matrix for MoveNet indicates high precision, recall, and F1 scores across all five yoga poses.

5.1 Comparative Study of both models:

Figures 12, 13 and 15 shows comparison of accuracies, weighted averages and macro averages of both the models respectively.

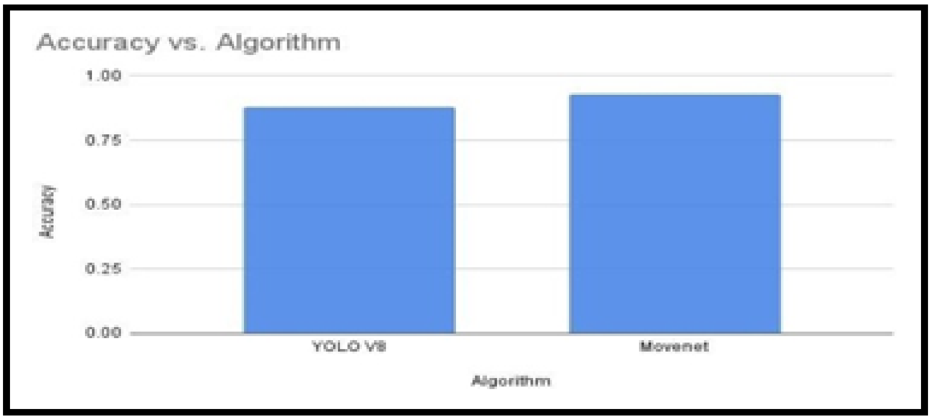


Fig. 12. Comparison of Accuracy of both models

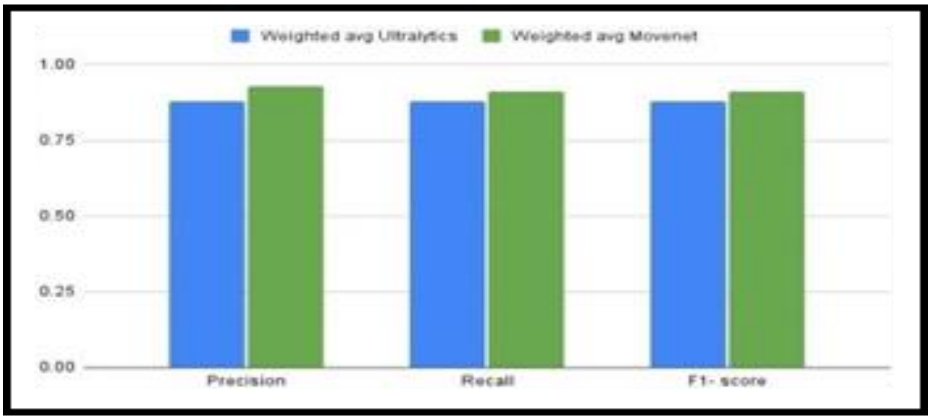


Fig. 13. Comparison of weighted averages

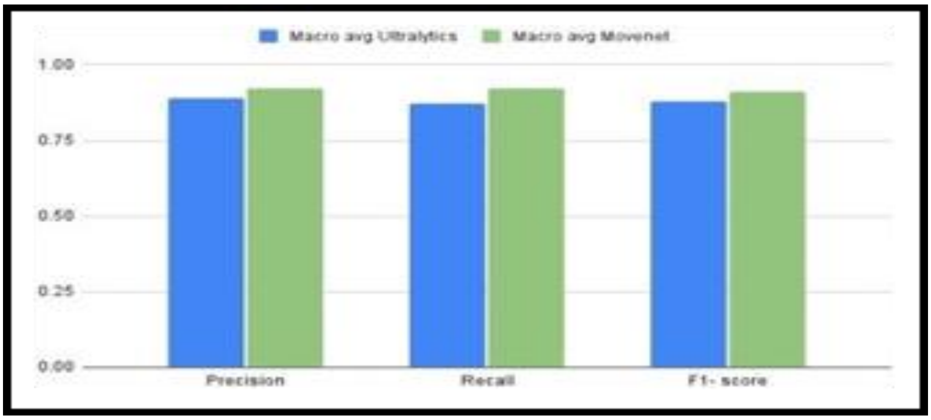


Fig.14. Comparison of macro averages

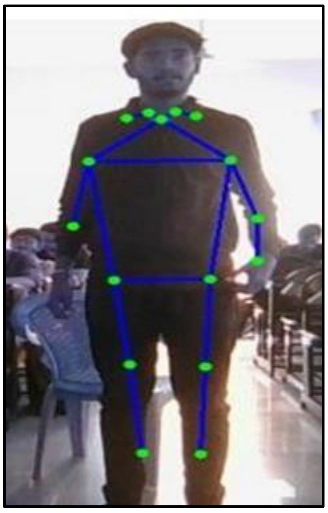
As it is evident from the diagrams, in summary, MoveNet outperforms YOLO-v8 making it superior model for recognizing yoga poses.

5.2 User Experience:

Users felt that the model is quite accurate while performing basic and intermediate poses with the responsiveness in real-time. They appreciated the user- friendly design and found it beneficial for personal yoga practice and remote coaching.

5.3 Results for Human Pose Estimation:

Figures 15 and 16 show real time pose estimation.



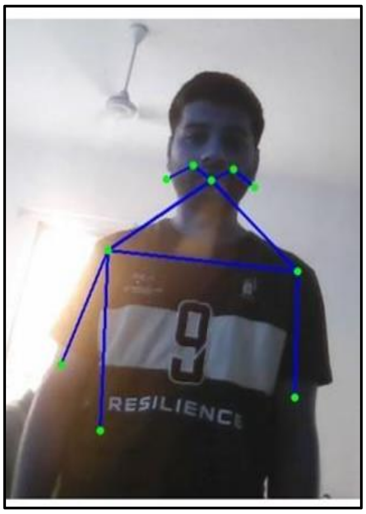


Fig. 15. Normal Standing pose **Fig.16.** Normal Standing pose

6. Conclusion

In conclusion, our research on human posture prediction with pre-trained models such as MoveNet and YOLOv8 and a yoga pose dataset has produced promising results. After careful examination, MoveNet has become the leader with an amazing accuracy of 93%, outperforming YOLOv8's 88% performance. This sharp difference highlights the superiority of specialised models like MoveNet in this domain and highlights how well they can capture yoga poses. The success of MoveNet and related pre-trained models represents a major advancement in the fields of yoga training, fitness tracking, and other related fields. These models provide fast development processes together with strong performance, which opens doors towards various real-world applications requiring accuracy and efficiency in human pose estimation. Our MoveNet-based real-time human position estimation technology provides notable improvements in speed and accuracy, allowing for a smooth integration into interactive applications like online yoga classes and fitness coaching. The MoveNet model has 93% accuracy which is better than YOLOv8 (88%). It works in real time and is mainly intended for yoga workout purposes while nowadays systems such OpenPose are more general, require more resources and can’t be as precise in detecting the yoga poses. The attained real-time performance satisfies the prerequisites for realistic implementation in real- world situations, opening the door for improved user experiences and novel paradigms for human- computer interaction. In future, to improve the model’s overall performance and reliability in real-world scenarios, we can strengthen their capacity to generalise over a range of poses and environmental variables through augmentation tactics like rotation, scaling, and noise injection. We acknowledge that offering consumers practical feedback while they execute yoga postures is just as important as accurately estimating poses.

References

1. Desai, Miral M. and Hiren K. Mewada. “A novel approach for yoga pose estimation based on in-depth analysis of human body joint detection accuracy.” PeerJ Computer Science 9 (2023): n. Pag.
2. Fard, Marjan Ghobadi, Gholam Ali Montazer and Davar Giveki. “A Novel Fuzzy Logic-based Method for Modeling and Recognizing Yoga Pose.” 2023 9th International Conference on Web Research (ICWR) (2023): 1-6.
3. Kotak, Devangi L., Jay Desai, Rishi Shah, Parth Goel and Krishna Patel. “Yoga Pose Classification using Angle Heuristic Approach.” 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA) (2022): 1709-1714.
4. Yadav, Santosh Kumar, Apurv Shukla, Kamlesh Tiwari, Hari Mohan Pandey and Shaik Ali Akbar. “An Efficient Deep Convolutional Neural Network Model For Yoga Pose Recognition Using Single Images.” ArXiv abs/2306.15768 (2023): n. Pag.

5. Nagargoje, Shrinivas, Adesh Santosh Shinde, Pranav Tapadiya, Om Shinde and Prof. Anita Devkar. “Yoga Pose Detection.” International Journal for Research in Applied Science and Engineering Technology (2023): n. Pag.
6. ATALAY AYDIN, Vildan. “Comparison of CNN-Based Methods for Yoga Pose Classification.” Turkish Journal of Engineering (2023): n. Pag.
7. Deepa, D., R. Velumani and S. Selvaraj. “Yoga Pose Estimation along with Human Posture Detection using Deep Learning Approach.” Applied and Computational Engineering (2023): n. Pag.
8. Gajbhiye, Rutuja, Snehal Jarag, Pooja Gaikwad and Shweta Koparde. “AI Human Pose Estimation: Yoga Pose Detection and Correction.” (2022).
9. Sunney, Jothika, Musfira Jilani, Pramod Pathak and Paul Stynes. ―A Real-time Machine Learning Framework for Smart Home-based Yoga Teaching System.‖ 2023 7th International Conference on Machine Vision and Information Technology (CMVIT) (2023): 107-114.
10. Gupta, Abhay, Kuldeep Gupta, K. N. Mahalakshmi Gupta and Kapil O. Gupta. “Human Activity Recognition Using Pose Estimation and Machine Learning Algorithm.” International Symposium on Intelligent Control (2021).
11. Verma, Manisha, Sudhakar Kumawat, Yuta Nakashima and Shanmuganathan Raman. “Yoga-82: A New Dataset for Fine-grained Classification of Human Poses.” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2020): 4472-4479.
12. Prakash, Madhura, Aishwarya. S, Disha Maru, N. Vishal Chandra and Varshini V. ―Yoga Posture Classification using Computer Vision.‖ International Journal of Engineering and Management Research (2021): n. pag.
13. Kishore, DMohan, S. Bindu and NandiKrishnamurthy Manjunath. “Estimation of Yoga Postures Using Machine Learning Techniques.” International Journal of Yoga 15 (2022): 137 - 143.
14. Trejo, Edwin W. and Peijiang Yuan. ―Recognition of Yoga poses through an interactive system with Kinect based on confidence value.‖ 2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM) (2018): 606-611.
15. Coskun, Huseyin. ―Human Pose Estimation with CNNs and LSTMs.‖ (2016).
16. Dataset- https://www.kaggle.com/datasets/niharika41298/yoga-poses-dataset