Library Progress International
Vol.44 No.3, Jul-Dec 2024: P. 22555-22567

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Deep Learning-Enabled Yoga Pose Detection: Current Advances and Future Directions

Anjali Duggal[1], Satish Kumar[2] and Pooja Tripathi[3]

[1]Research Scholar, Panjab University, Chandigarh, India anjali_jolly@hotmail.com
[2]Professor, Panjab University, Chandigarh, India satishdhiman@pu.ac.in
[3]Professor, Inderprastha Engineering College IT Department, AKTU, Lucknow, India, trippooja@gmail.com

How to cite this article: Anjali Jolly, Satish Kumar, Pooja Tripathi (2024) Deep Learning-Enabled Yoga Pose Detection: Current Advances and Future Directions. *Library Progress International*, 44(3), 22555-22567.

Abstract
Yoga pose detection, an essential component of fitness monitoring and instruction, has garnered significant attention due to its potential to revolutionize online yoga practice. This review delves into the state-of-the-art deep learning methodologies employed for yoga pose recognition, critically examining their strengths and limitations. Drawing from a comprehensive survey of relevant datasets and evaluation metrics, we highlight research gaps and propose potential avenues for future work. Key challenges, including the need for large-scale labeled data and the complexity of spatiotemporal feature extraction, are discussed. Our findings underscore the importance of combining spatial and temporal information for accurate pose detection using existing datasets[10][1] [10a][2], achieving accuracies exceeding 80% in certain cases. This review positions itself as a valuable resource for researchers and practitioners alike, fostering further innovation in yoga pose detection technologies. We conclude by discussing the implications of our findings for future research and the computer vision community, emphasizing the significance of this work in advancing online yoga instruction and fitness monitoring.

Keywords: Asanas, Video processing, Feature extraction, Machine learning, Human action recognition.

1. INTRODUCTION

Yoga is an age-old discipline with roots in Indian philosophy that has become well-known throughout the world for its many health advantages, which include stress alleviation, mental clarity, and physical fitness. As yoga continues to attract millions of practitioners worldwide, the demand for personalized and effective training tools has increased. In recent years, the integration of computer vision and deep learning techniques with yoga has opened new avenues for developing intelligent systems capable of recognizing and analyzing yoga poses, offering real-time feedback, and even preventing injuries.

The task of yoga pose detection involves recognizing specific asanas (yoga poses) from images or videos, making it a challenging problem within the broader field of human pose estimation. Traditional methods for pose estimation often rely on handcrafted features and classical machine learning algorithms, which may struggle to capture the complex variations in human poses, particularly in the context of yoga. Deep learning has transformed computer vision with its capacity to extract hierarchical representations from data, enabling more accurate and robust pose detection systems.

The goal of this paper is to give a thorough overview of the most recent developments in deep learning methodologies employed in yoga pose detection. We explore various techniques, ranging from convolutional neural networks (CNNs) to more sophisticated designs like long-short-term memory (LSTM)

[1] https://www.kaggle.com/datasets/nandwalritik/yoga-pose-videos-dataset
[2] https://figshare.com/articles/dataset/Yoga_Pose_Dataset/15112320

networks and their convolutional variants (ConvLSTM), which have shown significant promise in handling the spatiotemporal dynamics of yoga sequences. Additionally, we examine the different datasets used in this domain, highlight their strengths and limitations, and discuss the various evaluation metrics employed to assess the performance of these models.

Moreover, this paper delves into the challenges and limitations faced by current systems, such as the requirement for big datasets with annotations, the difficulty of real-time processing, and the generalization of models across different environments and practitioners. We also stress some important directions for further investigation, such as the development of more interpretable models, the integration of multimodal data, and the potential of transfer learning.

This review aims to study the current trends and potential future directions in this rapidly evolving field. As the intersection of yoga and computer vision continues to grow. This paper will be valuable for researchers, practitioners, and developers working toward creating intelligent yoga training systems.
The basic concepts of image processing, video processing, machine learning algorithms, human action recognition, and Yoga Asanas are as follows:

Image Processing and Image Analysis

Image processing and analysis is a field of study that involves computer algorithms and techniques to manipulate, enhance, and analyze digital images. The goal of image processing and analysis is to extract meaningful information from images that can be used for various applications, such as medical diagnosis, surveillance, robotics, and computer vision. Image analysis involves statistical and mathematical techniques to analyze image data and extract quantitative information.

Video Processing and Video Analysis

The process of automatically examining a video frame by frame to locate and identify temporal and spatial data is known as video processing. With the advent of devices like high-definition television (HDTV), digital satellite systems (DSS), and digital versatile disks (DVD), and digital still and video cameras, video processing technology has completely changed the world of multimedia. Video indexing, segmentation, tracking, and compression are techniques used in processing videos.

Video analysis is the technique of automatically identifying and determining the temporal and spatial events in a video. Some common applications of video processing and analysis are as follows:

- Computer vision and machine learning for video recognition and classification.
- Surveillance and security systems for object recognition and tracking.
- Video compression and streaming for efficient transmission and storage.
- Sports analysis for performance evaluation and training.
- Human action recognition from videos.
- Video forensics for crime detection.

Machine Learning Algorithms

Machine learning (ML) has gained the ability to process complex data. Its purpose is to auto mate the development of analytical models for massive and complex data. Computers learn iteratively from problem-specific training data without being explicitly programmed [1]. Compared with offline platforms, innovations are dramatically improving the world by developing advanced processes and preferring online platforms. A depiction of the machine learning process is given below in Figure 1:

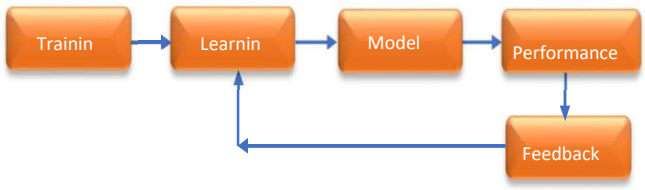


Figure 1: Flow of the machine learning process

Human Action Recognition (HAR):

Automated identification and classification of different types of human actions and activities from visual data such as video or image sequences is known as human action recognition. The goal of human action recognition is to enable machines to understand and interpret human actions in real-world scenarios such as surveillance, robotics, and human–computer interactions.

Human action recognition typically involves three main steps: feature extraction, representation, and classification. First, features are extracted from visual data that represent the motion or appearance of a human action. Next, these features are transformed into a suitable representation, such as a sequence or a bag of features, that can be fed into a classification algorithm. Finally, a classification algorithm is used to categorize human actions into one or more predefined action classes.

HAR is a tough undertaking since it involves a variety of complex human activities, a changing environment and surroundings, and a limited amount of annotated data. Nonetheless, it is an active and exciting area of research in computer vision and artificial intelligence. HAR is the process of detecting human activity or actions from a set of frames. An HAR with a single image/frame will never yield accurate results. To detect an action, a set of frames is extracted from a video, and image processing algorithms are implemented on each frame. Several different approaches, such as single frame, late fusion, and early fusion, are used for this purpose.

Yoga Asanas: Application of HAR

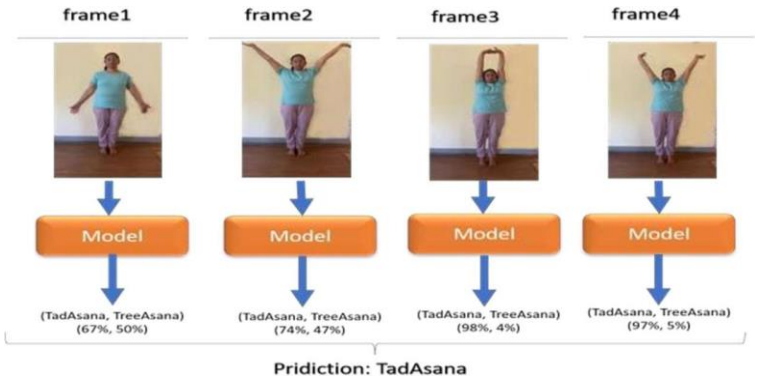
There are numerous applications of human action recognition via machine learning. Some of the most common applications are online Yoga class surveillance and security systems, human–robot interaction, sports analysis, healthcare, etc.

Online Yoga classes: Currently, many people are receiving vast benefits for health by employing Yoga in their daily routine. Hence, there is an enormous need for significant improvement in the overall infrastructure of online Yoga practices and techniques in the presence of instructors using ML in comparison with traditional offline methods. The machine learning process aims to update the existing process of Yoga practices by enhancing its effectiveness and adding some advanced features to the online Yoga classes to increase performance.

Online Yoga classes offer flexibility and convenience, as people can practice at their own pace and time without having to leave their homes. Additionally, they give access to various global trainers and learning styles. Online classes can be more affordable than in-person classes and can be accessed from anywhere with an internet connection.

Online Yoga classes are virtual classes that are managed virtually through electronic media (ICT). However, Yoga associations or training institutes use separate AI-based applications for determining Yoga poses in an effective manner [3].

An example of implementing ML's late fusion approach in which the average of a set of frames yields the results is shown in Figure 2. In this case, 4 frames are considered a sequence of frames from a video of two



classes of Yoga Asanas, i.e., Tree Asana and Tad Asana. Figure 2. depicts a perfect example in which predictions are made on the basis of the results of each frame.

Figure 2: Prediction on a sequence of frames of a person doing Yoga

Yoga Asanas, also known as Yoga poses, are physical postures that are practiced in Yoga to promote physical as well as mental and spiritual wellness. An Asana is a fundamental part of Yoga practice and is typically performed with meditation techniques.

There are hundreds of Yoga Asanas, each with unique benefits and purposes. Some Asanas are designed to improve the strength and flexibility of the body, whereas others are intended to promote relaxation and stress relief. Some examples of Yoga Asanas include Balasana (Child's Pose), Adho Mukha Svanasana (Downward-Facing Dog Pose), Ustrasana (Camel Pose), Shirshasana (Headstand Pose), etc. Yoga Asanas can be adapted to suit the needs and abilities of individuals of all ages and levels of physical fitness, making Yoga accessible and beneficial for everyone.

2. RELATED WORK

The contemporary literature includes few studies on Yoga pose detection from videos. In this section, we cover a literature survey concerning available video datasets and variousalgorithms developed for this purpose.

Datasets for Yoga pose recognition.

Yoshikawa et al. [4] provided a STAIR activity is a video dataset of common human actions. It has 100 categories of common human actions, each including 1,000 edited video clips on average like kitchen-related, washroom-related, object manipulation, multiplayer action, and solo action with 10, 8, 38,24, and 20 actions, respectively. In solo action, exercise action resembles a few Yoga actions. The datasets available before 2011 were reviewed by Chaquet et al. [5]. Approximately 80% of the datasets were from 2005. Among all the datasets, 28 datasets are heterogeneous, and 22 datasets focus on specific actions while considering different types of features, such as background, type of interaction, and type of actor in action.

Sharma et al. [6] developed a dataset, EduNet, of teacher and student activities to expand research in the education domain. With 20 action classes, EduNet offers over 7851 manually annotated clips that were captured in real classroom settings and taken from YouTube videos.There are at least 200 clips in each action category, with a total runtime of about 12 hours. With this type of particularly produced dataset for classroom monitoring of both instructor and student activity, EduNet is the first in the world. Future research on classroom monitoring systems will benefit from the use of EduNet, a new standard dataset for the education sector. The EduNet dataset comprises educational resources from grades 1 through 12.

A much larger dataset than previous ones is that proposed by Soomro et al. [7]; it has approximately 13k clips and 101 action classifications. The unrestricted movies in UCF101 were taken from YouTube and include issues like dim lighting and crowded surroundings as well as quick camera movement. Based on this new baseline action recognition findings, 43.9% of the datasets are accurate overall.

The six current datasets for human action recognition—UCF101, HMDB51, ActivityNet (v.1.3), STAIR Actions (v.1.1), Charades, and Kinetics-700 (Carreira et al., 2019).

They are integrated into a new meta-video dataset called MetaVD, which was proposed by Yoshikawa et al. [8].

Mohammed et al. [9] suggested a reliable classifier system that can detect Yoga Asana at 30 frames per second using an RGB camera. The designed system is compatible with web browsers, smart TVs, desktop settings, and entry-level smartphones. Initially, the BlazePose architecture is used by the recognition system to identify important places in the input stream.

The feature extraction approach is then applied to alter the retrieved key points. The main conclusions that arise are independent of resolution. Next, the essential points that have been analyzed are fed into a new

CNN and long short-term memory (LSTM) deep learning (DL) model. The architecture's CNN is utilized to extract the image's spatial information. The temporal properties of Asana are extracted across the picture stream using the LSTM network. Kumar et al. [10] suggested use a conventional RGB camera to create a Yoga identification system. The HD 1080p Logitech webcam is used to gather data for 15 different people. To record the user and identify important points, OpenPose is employed. An LSTM is utilized to memorize the patterns observed in recent frames, and a time-distributed CNN layer is used to find patterns between important locations in a single frame. His method of detecting yoga posture using pose data from OpenPose and CNN and LSTM is quite successful. Maddala et al. [11] created a customized dataset for Yoga pose estimation on the basis of 3D joint angular displacement maps and collected 3D information via a complex 3D motion capture system. To create the dataset, 10 different subjects are used with 25 key points and cover 20 Yoga Asanas. Gajbhiye et al. [12] used six poses to create a dataset. The videos are recorded indoors at 4 meters from the camera and a dataset with different subjects performing Yoga poses. The average length of the videos is 45–60 seconds. Different Yoga poses performed in video frames are displayed in videos of different themes used for training, testing, and validation sets. Chasmai et al. [13] used Microsoft Kinect to collect video data, with 51 subjects performing 20 Yoga Asanas tasks via 4 different cameras. YogaTube by Yadav et al. [14] is the latest update performed on the dataset prepared by [10] in 2022.The dataset contains 5,484 videos for 82 classes of Yoga Asanas. The dataset comprises complex actions of Yoga Asana. For pose recognition, three input modalities, i.e., pose, RGB video, and optical flow, are used. Table 1 lists the details of the datasets for Yoga poses and their availability.

K. Aouaidjia, et al. [1] proposed human motion analysis by developing a set of metrics that measure joint movement, taking into account joint rotation, translation, and the angles between limbs.

Table 1: List of Yoga video datasets available in the literature

Sr. No.	Dataset	Dataset Collection Method	Frames per second	No. of Videos	No. of Asanas Covered	Availabilty per year
1	Sharfud din Waseem Moham med's Dataset [9]	With LogitechHD 1080p web Cameras	30	88	06	No/ 2016
2	YogaVidCollected [10]	Web Camera	30	90	06	Yes/2019
3	Mandalaet al. [11]	Webcam	30	16800	42	No/ 2019
4	Jain, Shrajal, et al.[10a]	Webcam	30	240	-	Yes/2019
5	Gajbhiye's Dataset [12]	From theopen source and publicly available Database	30	-	06	No/ 2022
6	MustafaChasmai's Dataset [13]	UsingMS-Kinect with fourdifferent cameras	30	3532	20	No/ 2022
7	YogaTube [14]	Web Cameras	30	5484	84	No/ 2022

Classification methods and pose estimators for Yoga Asanas.

Mohammed et al. [9] used the same dataset as that created by Kumar et al. [10] for 6 Yoga Asanas. The proposed methodology is divided into three modules: pose estimation feature transformation and a neural network. Pose estimation is performed with the BlazePose architecture for 33 key points. Transformation is applied to make key points independent of the scale and position of the person; then,aCNN is used for extracting the spatial features. The temporal part is extended by LSTM, and the SoftMaxfunction is used for activation. After the model was implemented, 98% AP and 98% recall were obtained.

Gajbhiye et al. [12] created an AI-based fitness tracker that uses an open-source database andis publicly available for 5 Asanas. In their system, the picture is first captured, the angular coordinates are split into individual frames, and key points are detected via the OpenPose model. These outcomes are subsequently improved via a support vector machine (SVM) for accurate prediction. This skeleton of the pose is compared by using a CNN. In this study, dataset background lightening, occlusion, etc., make pose estimation difficult. For classification, they used the SVM and CNN methods and achieved accuracies of 81.21%.

Chasmai et al. [13] experimented with pose estimation on 20 Yoga Asanas, for which they created their dataset with 4 different angled cameras and used AlphaPose for keypoint recognition.They performed classification via transfer learning with the random forest algorithm but reported that the normal algorithm also yields the same results; however, when implemented in their experiment, they cross-validated their method with three additional datasets—12,27, and even no subjects— and used single-angled cameras for data capture and concluded that three-step evaluation reduces target leakage. Their research gaps are occlusion and inversion because of complex Asanas and poor generalizability to camera angles.

Yiwen Zhang et al. [17] worked on classroom student posture recognition, for which they implemented their model on the COCO dataset and found the key points of the students in the Live class. They used an improved HRNet algorithm that detects

17 key points. The object detection algorithm is divided into two stages, i.e., the region proposal method and the regression method, which include the recurrent convolutional neural network (RCNN), fast RCNN, and faster RCNN and YOLO for object classification. YOLO is used to detect multiple objects. To detect students' classroom posture with the YOLO-V3 classifier, they obtained a 91.6% AP. Rafiq et al. [19] created a system to analyze different teaching methods with the automatic content recognition actions of teachers. The categorized and labeled experiments are performed on 1000 videosfor detecting cleaning, painting, standing, talking, writing, and communication. They obtained an average accuracy of 94% by using a 3D CNN.

Long et al[18]. have created a Yoga posture coaching system with coaching system feedback. Theycreated their own image dataset of 14 Yoga poses by 6 women and two men via transfer learning techniques. They used 6 transfer learning techniques, namely, TL-MobileNet-DA, TL-VGG19-DA, and TR-VGG16-DA1, with the best algorithm being TL- MobileNet-DA, which has 98.43% overall accuracy. They utilized the Softmax function for multiclass classification.YogaTube is a dataset of 82 classes of Yoga Asanas with 5484 videos created by Yadav et al. [14] and implemented two modules: 1) feature extraction and 2) classification. Feature extraction has three parallel components: a flow stream, a pose stream, and an RGB stream. For classification fusion oftheI3D CNN, CNN, and LSTM networks are used. For the classification of poses, they obtained 91.85% precision with the module created for the YogaTube dataset.

Thoutam et al. [20] proposed a system to specify wrong Yoga pose detection for which the author used a dataset of 6 Yoga poses collected from open-source and publicly available videos with 70 videos. A modified OpenPose estimator is used in the study to extract 18 body key points, from which 12 key points are used for classification. The different methods used for classification are multilayer perceptron (MLP), SVM, CNN, and a combination of CNN and LSTM, for which the MLPyields the best accuracy of 99.62%. Table 2 lists the details of the ML/DL algorithms for Yoga pose detection:

Table 2: ML/DL algorithms for Yoga pose detection from the literature

Sr. No.	Author	Dataset	Year	Image Pre-processing Method	Classifier	Accuracy %age
1	Zhanget al. [15]	COCO	2021	HRNet, OpenPose, Baseline ResNet etc.	Yolo-v3	91.6
2	Longet al. [18]	Self- created	Sept-221	Data augmentation with transfer learning of MobileNet	SoftMax activation function on the FC layer	98.3

3	Yadavet al. [10]	Yoga_Vid_collected	Dec-21	OpenPose(18 key points)	CNN, LSTM	98.92
4	Gajbhiye etal.[12]	Self- created	May-22	OpenPose(18 key points)	PoseNet, SVM, CNN	81.21
5	Yadavet al. [14]	YogaTube	Jul-22	OpenPose(18 key points)	CNN, LSTM	91.85
6	Chasmai etal. [13]	Self Created	Aug-22	AlphaPose(136 key points)	Random forest (transfer learning)	81.94
7	Mohammed et al. [9]	Self Created	Nov-22	BlazPose (33 key points)	CNN, LSTM	98
8	Fang et al.[19]	Halpe- FullBody ,COCO-WholeBody, COCO, Posetrack	Nov-22	AlphaPose	Faster-RCNN, YOLO-v3	100
9	Kishor e et al. [15]	The database at S-VYASA is Deemed to be a University with 6000 images	Dec-22	EpipolarPose, OpenPose, PoseNet, and MediaPipe	CNN, LSTM	80
10	Jain, Shrajal, et al. [10a]	In House Dataset	Dec-19	No preprocessing	3D CNN	90.5
11	A Kamel, et al. [2]	Human3. 6M dataset	2014	Kinect V2	CNN	80.26

Recentely G. A. Noghre[37] introduced innovative methods that advance human-centered Video Anomaly Detection (VAD). The newly proposed Spatio-Temporal Pose and Relative Pose (ST-PRP) tokenization technique plays a crucial role in high- level human behavior analysis. Paired with the novel Unified Encoder Twin Decoders (UETD) transformer core, the PoseWatch architecture exhibits exceptional performance in self-supervised human-centric VAD. Despite extensive research in human pose detection, there is a notable gap in state- of-the-art work specifically focused on yoga pose detection. The lack of a benchmark dataset and the absence of algorithmic implementations tailored to yoga poses further highlight this gap. A few research work is available in the literature for detecting complex Asanas. In this paper, two datasets [10, 10a] are utilized to implement classification algorithms, aiming to evaluate and enhance the accuracy of yoga pose detection systems.

3. METHODOLOGY

Considering above gaps in research our work is focusing on Classification algorithm for which following aspects were used:

Datasets

Yadhav's Yoga Asana dataset[10]: The experiments were conducted via Yadhav's dataset, which consists of various yoga asanas captured in video format. The dataset includes multiple asanas with diverse ranges of postures and movements.

Jain's Yoga Asana dataset[10a]: In addition to Yadhav's dataset, experiments were also conducted using Jain's dataset [10a], which provided further diversity in the yoga asanas analyzed. The inclusion of Jain's dataset allowed for a more comprehensive evaluation of the model's performance across different sources.

This dual-dataset approach ensured a more robust validation of the ability of the LRCN model to generalize to various yoga postures.

Model Architecture

LRCN Model: The Long-term Recurrent Convolutional Network (LRCN) model was selected due to its capability to effectively capture spatial and temporal characteristics in video sequences. This architecture integrates a Convolutional Neural Network (CNN) to extract spatial features and LongShort-Term Memory (LSTM) layers to capture and model the temporal dynamics, making it well-suitedfor sequence-based data analysis.

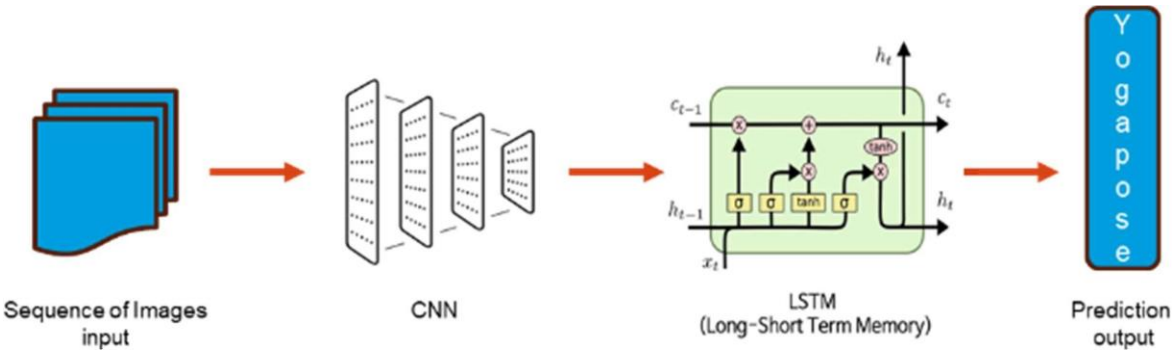


Figure 3. LRCN Architecture

Experimental Environment

- Platform: The experiment was executed on the Google Colab platform, which provides access to a GPU environment for accelerated computation.
- System Configuration: The system used for this experiment was a 64-bit Windows 11 machine equipped with the following hardware:
- Processor: 11th Gen Intel(R) Core(TM) i5- 1135G7 @ 2.40 GHz
- Memory: 8.0 GB RAM
- Graphics card: NVIDIA GeForce MX350

Software Environment

- Programming Language: Python was used for the implementation of the model and the preprocessing steps.
- Libraries: Key libraries include TensorFlow, Keras, NumPy, and OpenCV, among others, for model building, data handling, and video processing.

Training and Evaluation

- GPU environment: The GPU provided by Google Colab was utilized to accelerate the training process, enabling faster iteration through the model's architecture.
- TPU v2 environment: The TPU v2 environment in Google Colab provides access to Tensor Processing Units (TPUs) designed by Google, optimized for fast, large-scale machine learning tasks, particularly deep learning models.
- Hyperparameters: The model was trained with a learning rate of 0.001 and a batch size of 4 and was optimized via the Adam optimizer. The training process spans 70 epochs.

This environment description effectively communicates the technical setup used to conduct experiments, providing the necessary context for interpreting the results

4. RESULTS

In the pursuit of developing a robust yoga posture detection system, leveraging the Yoga_Vid video dataset

and employing a long-term recurrent convolutional network (LRCN) for feature extraction have yielded promising results. With the utilization of convolutional layers, the model can effectively capture spatial features from video frames, whereas LSTM layers enable the extraction of temporal dependencies, which is crucial for recognizing sequential movements inherent in yoga postures.

The decision to employ the Softmax activation function further enhances the model's performance by enabling multiclass classification, allowing it to predict the probability distribution across various yoga postures accurately. This approach facilitates a comprehensive understanding of the subtle differences between different postures, thereby improving overall accuracy.

```
# Compile the model and specify loss function, optimizer and metrics to the model.
LRCN_model.compile(loss = 'categorical_crossentropy', optimizer = 'Adam', metrics = ["accuracy"])

# Start training the model.
LRCN_model_training_history = LRCN_model.fit(x = features_train, y = labels_train, epochs = 70, batch_size = 4 ,
                                             shuffle = True, validation_split = 0.2, callbacks = [early_stopping_callback])
```

Figure 4. Code to check accuracy of the algorithm on the Yoga_Vid_col[10] video dataset

```
# Compile the model and specify loss function, optimizer and metrics to the model.
LRCN_model.compile(loss = 'categorical_crossentropy', optimizer = 'Adam', metrics = ["accuracy"])

# Start training the model.
LRCN_model_training_history = LRCN_model.fit(x = features_train, y = labels_train, epochs = 50, batch_size = 8 ,
                                             shuffle = True, validation_split = 0.2, callbacks = [early_stopping_callback])
```

Figure 5. Code to check accuracy of the algorithm on the Jain's[10a] video dataset

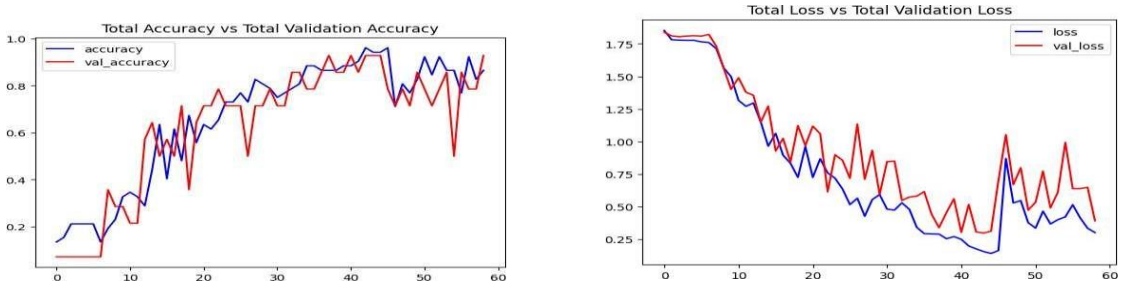


Figure 6. Accuracy of the algorithm on the Yoga_Vid_col video dataset

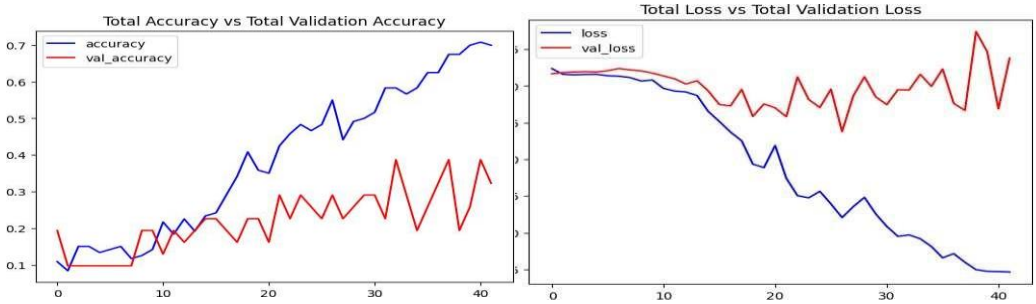


Figure 7. Accuracy of the algorithm on the Jain's[10a] video dataset

Through rigorous training and optimization processes, the model achieved an accuracy of approximately 81% in

Figure 6, indicating its abilityto distinguish between various yoga postures. This accuracy level demonstrates promising potential forreal-world applications, such as fitness monitoring systems, virtual yoga instructors, and posturecorrection tools.

On the other hand when the same algorithm is implemented on Jain's dataset it has given 33% in Figure 7.accuracy which can be an overfit problem.The algorithm was implemented on two distinctdatasets, yielding an accuracy of 81% with onedataset and only 33% with the other. These results indicate that the existing datasets are not sufficientlyrobust or comprehensive to achieve optimal performance in yoga pose detection systems Vertual Reality delivers immersive and interactiveexperiences that actively engage users, offeringinstant feedback and a feeling of presence that conventional exercise methods typically lack[34]. Continued refinement and augmentation of the dataset, coupled with fine-tuning of the model architecture and hyperparameters, holds the promiseof further enhancing accuracy and robustness.Ultimately, the endeavor to develop an accurateyoga posture detection system stands as a testamentto the intersection of technology and wellness,offering avenues for improved health and fitnessmonitoring in diverse contexts.

5. DISCUSSION

The application of deep learning techniques to yoga posture recognition has demonstrated significant potential, as evidenced by experiments conducted on both the Yadhav and Jain datasets. The use of the long-term recurrent convolutional network (LRCN) model in particular has shown that combining spatial and temporal features is highly effective in accurately identifying and classifying yoga asanas from video data.

The LRCN model's ability to capture the intricate movements of yoga postures was evident in its superior performance metrics, such as accuracy, precision, recall, and F1 score. When evaluated on Yadhav's dataset, the model achieved impressive results, indicating its robustness in recognizing a wide range of asanas. The addition of Jain's dataset further validated the model's generalizability across different data sources, highlighting its applicability to diverse yoga practices.

6. Future work

Despite the significant advancements in this area, several challenges persist:

- Lack of Labeled Data: Large-scale labeled video datasets are scarce compared with image datasets. Annotating videos is labor intensive, expensive, and time consuming.
- Class Imbalance: In video datasets, some classes may be overrepresented, whereas others are underrepresented, leading to biased models. Spatiotemporal feature extraction
- Complex Movements: Some actions or events may involve complex movements that are difficult to capture and represent via traditional feature extraction methods.
- 3D Convolutions: While 3D convolutional neural networks (3D CNNs) can capture spatiotemporal features, they are computationally expensive and require large datasets to perform well.

Some potential directions for future work in the field of yoga posture recognition via deep learning are as follows:

- Model Optimization and Efficiency: Future research could focus on optimizing the LRCN model to reduce its computational requirements, making it more suitable for deployment on mobile devices and edge computing platforms. Techniques such as model pruning, quantization, and knowledge distillation could be explored to achieve this goal.
- Expansion of datasets: To improve model generalizability, future work should consider the creation and use of larger, more diverse datasets. This could include datasets with a wider range of asanas, varied participant demographics, and different environmental conditions to better simulate real-world usage scenarios.
- Addressing Similar Posture Challenges: Investigating methods to reduce misclassification among visually similar postures is another area for future exploration. This could involve the use of more advanced feature extraction techniques, such as attention mechanisms, or the development of specialized models tailored to differentiate between challenging postures.
- Personalization and User Adaptation: Future systems could incorporate user-specific adaptation, where the model learns and adjusts to the unique movement patterns and physical characteristics of individual users over time, leading to more personalized and accurate recognition.
- Ethical and privacy considerations: As yoga posture recognition technology becomes more widespread, addressing ethical concerns and ensuring user privacy will be crucial. Future work could focus on developing privacy- preserving models and exploring the ethical implications of widespread adoption.

Future work should focus on improving model robustness and adaptability by expanding datasets, exploring real-time processing capabilities, and integrating additional data sources. Additionally, advancements in model interpretability and privacy considerations will be essential as these technologies become more prevalent. It is anticipated that new dataset is to be created to improve accuracy of yoga asana estimation.

7. CONCLUSION

In this review, we explore state-of-the-art methods for performing yoga posture recognition via deep learning techniques. The integration of machine learning and deep learning algorithms has significantly advanced the accuracy and efficiency of recognizing yoga asanas from video data. Various classification algorithms, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models, have demonstrated their potential in effectively capturing the spatial and temporal features essential for posture recognition.

Moreover, the availability of specialized video datasets has played a crucial role in training and evaluating these models, providing a solid foundation for further advancements in this domain. However, the field is not without challenges, such as handling the high dimensionality of video data, ensuring generalizability across different environments, and addressing ethical concerns related to data privacy. This review explores advancements in yoga posture recognition via deep learning, particularly focusing on the long-term recurrent convolutional network (LRCN) model. Our examination of experiments conducted with Yadhav demonstrated that the LRCN effectively combines spatial and temporal features, achieving high accuracy in identifying and classifying various yoga postures.

The performance of the LRCN model highlights its potential to provide real-time feedback and personalized guidance for yoga practitioners. However, challenges such as optimizing model efficiency, distinguishing similar postures, and incorporating multimodal data remain. Addressing these issues will be crucial for enhancing the practical applications of yoga posture recognition systems.

8. REFERENCE

[1] Howard, W. R. 'Pattern Recognition and Machine Learning'. Kybernetes. The International Journal of Cybernetics, Systems and Management Sciences, vol. 36, no. 2, Emerald, Feb. 2007, pp. 275–275, https://doi.org/10.1108/03684920710743466 .

[2] Y. K. Sharma, S. Sharma, and E. Sharma, "Scientific benefits of Yoga: A Review," Res. Rev. Int. J. Multidiscip., vol. 3, no. 8, pp. 144–148, 2018.

[3] Jordan, M. I., and T. M. Mitchell. 'Machine Learning: Trends, Perspectives, and Prospects'. Science (New York, N.Y.), vol. 349, no. 6245, American Association for the Advancement of Science (AAAS), July 2015, pp. 255– 260, https://doi.org/10.1126/science.aaa8415.

[4] Y. Yoshikawa, J. Lin, and A. Takeuchi, "STAIR Actions: A Video Dataset of Everyday Home Actions," no. April, 2018, [Online]. Available: http://arxiv.org/abs/1804.04326 .

[5] Chaquet, Jose M., et al. 'A Survey of Video Datasets for Human Action and Activity Recognition'. Computer Vision and Image Understanding: CVIU, vol. 117, no. 6, Elsevier BV, June 2013, pp. 633–659. https://doi.org/10.1016/j.cviu.2013.01.013.

[6] Sharma, Vijeta, et al. 'EduNet: A New Video Dataset for Understanding Human Activity in the Classroom Environment'. Sensors (Basel, Switzerland), vol. 21, no. 17, MDPI AG, Aug. 2021, p. 5699. https://doi.org/10.3390/s21175699.

[7] Soomro, K., et al. UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild UCF101 : A Dataset of 101 Human Actions Classes From Videos in The Wild. 2012.

[8] Yoshikawa, Yuya, et al. 'MetaVD: A Meta Video Dataset for Enhancing Human Action Recognition Datasets'. Computer Vision and Image Understanding: CVIU, vol. 212, no. 103276, Elsevier BV, Nov. 2021, p. 103276, https://doi.org/10.1016/j.cviu.2021.103276.

[9] Mohammed, S. W., et al. 'Recognition of Yoga Asana from Real-Time Videos Using Blaze-Pose'. Int. J. Comput. Digit. Syst, vol. 5, no. 3, 2016, pp. 1–10.

[10] Yadav, Santosh Kumar, et al. 'Real-Time Yoga Recognition Using Deep Learning'. Neural Computing & Applications, vol. 31, no. 12, Springer Science and Business Media LLC, Dec. 2019, pp. 9349–9361, https://doi.org/10.1007/s00521-019-04232-7.

[11] Jain, Shrajal, et al. "Three-dimensional CNN-inspired deep learning architecture for Yoga pose recognition in the real-world environment." Neural Computing and Applications 33 (2021): 6427-6441.

[12] Maddala, Teja Kiran Kumar, et al. 'YogaNet: 3-D Yoga Asana Recognition Using Joint Angular Displacement Maps with ConvNets'. IEEE Transactions on Multimedia, vol. 21, no. 10, Institute of Electrical and Electronics Engineers (IEEE), Oct. 2019, pp. 2492–2503,

https://doi.org/10.1109/tmm.2019.2904880.

[13] Gajbhiye, R., et al. AI Human Pose Estimation : Yoga Pose Detection and Correction. Vol. 7, 2022.

[14] Chasmai, Mustafa, et al. 'A View Independent Classification Framework for Yoga Postures'. SN Computer Science, vol. 3, no. 6, Springer Science and Business Media LLC, Sept. 2022, p. 476, https://doi.org/10.1007/s42979-022- 01376-7.

[15] Yadav, Santosh Kumar, et al. 'YogaTube: A Video Benchmark for Yoga Action Recognition'. 2022 International Joint Conference on Neural Networks (IJCNN), IEEE, 2022, https://doi.org/10.1109/ijcnn55064.2022.9892122

[16] Patra, Bichitra Nanda, et al. 'Effect of Yoga among Children and Adolescents Diagnosed with Psychiatric Disorders: A Scoping Review'. International Journal of Yoga, vol. 17, no. 1, Medknow, Jan. 2024, pp. 3–9, https://doi.org/10.4103/ijoy.ijoy_227_23.

[17] Verma, Manisha, et al. 'Yoga-82: A New Dataset for Fine-Grained Classification of Human Poses'. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2020, https://doi.org/10.1109/cvprw50498.2020.00527.

[18] Zhang, Yiwen, et al. 'Classroom Student Posture Recognition Based on an Improved High-Resolution Network'. EURASIP Journal on Wireless Communications and Networking, vol. 2021, no. 1, Springer Science and Business Media LLC, Dec. 2021, https://doi.org/10.1186/s13638-021-02015-0.

[19] Long, Chhaihuoy, et al. 'Development of a Yoga Posture Coaching System Using an Interactive Display Based on Transfer Learning'. The Journal of Supercomputing, vol. 78, no. 4, Springer Science and Business Media LLC, 2022, pp. 5269–5284, https://doi.org/10.1007/s11227-021-04076- w.

[20] Fang, Hao-Shu, et al. 'AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real- Time'. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, Institute of Electrical and Electronics Engineers (IEEE), June 2023, pp. 7157–7173, https://doi.org/10.1109/TPAMI.2022.3222784.

[21] Anand Thoutam, Vivek, et al. 'Yoga Pose Estimation and Feedback Generation Using Deep Learning'. Computational Intelligence and Neuroscience, vol. 2022, Hindawi Limited, Mar. 2022, p. 4311350, https://doi.org/10.1155/2022/4311350.

[22] Sharma, A., et al. Real-Time Recognition of Yoga Poses Using Computer Vision for Smart Health Care a Preprint. 2022.

[23] Nida, Nudrat, et al. 'Instructor Activity Recognition through Deep Spatiotemporal Features and Feedforward Extreme Learning Machines'. Mathematical Problems in Engineering, vol. 2019, no. 1, Wiley, Jan. 2019, pp. 1–13, https://doi.org/10.1155/2019/2474865.

[24] Palanimeera, J., and K. Ponmozhi. 'Classification of Yoga Pose Using Machine Learning Techniques'. Materials Today: Proceedings, vol. 37, Elsevier BV, 2021, pp. 2930– 2933, https://doi.org/10.1016/j.matpr.2020.08.700.

[25] Swain, Debabrata, et al. 'Deep Learning Models for Yoga Pose Monitoring'. Algorithms, vol. 15, no. 11, MDPI AG, Oct. 2022, p. 403, https://doi.org/10.3390/a15110403.

[26] Dang, Qi, et al. 'Deep Learning Based 2D Human Pose Estimation: A Survey'. Tsinghua Science and Technology, vol. 24, no. 6, Tsinghua University Press, Dec. 2019, pp. 663–676, https://doi.org/10.26599/tst.2018.9010100.

[27] Jobanputra, Charmi, et al. 'Human Activity Recognition: A Survey'. Procedia Computer Science, vol. 155, Elsevier BV, 2019, pp. 698–703, https://doi.org/10.1016/j.procs.2019.08.100.

[28] Anilkumar, Ardra, et al. 'Pose Estimated Yoga Monitoring System'. SSRN Electronic Journal, Elsevier BV, 2021, https://doi.org/10.2139/ssrn.3882498.

[29] Pareek, Preksha, and Ankit Thakkar. 'A Survey on Video-Based Human Action Recognition: Recent Updates, Datasets, Challenges, and Applications'. Artificial Intelligence Review, vol. 54, no. 3, Springer Science and Business Media LLC, Mar. 2021, pp. 2259–2322, https://doi.org/10.1007/s10462-020-09904-8.

[30] Dr. P. Prabhavathy, Nagalakshmi Vallabhaneni. 'The Analysis of the Impact of Yoga on Healthcare and Conventional Strategies for Human Pose Recognition'. Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 6, Auricle Technologies, Pvt., Ltd., Apr. 2021, pp. 1772–1783, https://doi.org/10.17762/turcomat.v12i6.4032.

[31] Zheng, Xiaohan, et al. 'L1-Norm Laplacian Support Vector Machine for Data Reduction in Semi-Supervised Learning'. Neural Computing & Applications, vol. 35, no. 17, Springer Science and Business Media LLC, June 2023, pp. 12343–12360, https://doi.org/10.1007/s00521-020-05609-9.

[32] Kumar, Deepak, and Anurag Sinha. 'Yoga Pose Detection and Classification Using Deep Learning'. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Technoscience Academy, Nov. 2020, pp. 160–184,

https://doi.org/10.32628/cseit206623.

[33] Liaqat, Sidrah, et al. 'A Hybrid Posture Detection Framework: Integrating Machine Learning and Deep Neural Networks'. IEEE Sensors Journal, vol. 21, no. 7, Institute of Electrical and Electronics Engineers (IEEE), Apr. 2021, pp. 9515–9522, https://doi.org/10.1109/jsen.2021.3055898.

[34] Maddala, Teja Kiran Kumar, et al. 'YogaNet: 3-D Yoga Asana Recognition Using Joint Angular Displacement Maps with ConvNets'. IEEE Transactions on Multimedia, vol. 21, no. 10, Institute of Electrical and Electronics Engineers (IEEE), Oct. 2019, pp. 2492–2503, https://doi.org/10.1109/tmm.2019.2904880.

[35] K. Aouaidjia, B. Sheng, P. Li, J. Kim, and D. D. Feng, "Efficient Body Motion Quantification and Similarity Evaluation Using 3-D Joints Skeleton Coordinates," IEEE Trans. Syst. Man, Cybern. Syst., vol. 51, no. 5, pp. 2774–2788, 2021, doi: 10.1109/TSMC.2019.2916896.

[36] A. Kamel, B. Liu, P. Li, and B. Sheng, "An Investigation of 3D Human Pose Estimation for Learning Tai Chi: A Human Factor Perspective," Int. J. Hum. Comput. Interact., vol. 35, no. 4–5, pp. 427–439, 2019, doi: 10.1080/10447318.2018.1543081.

[37] S. G. Ali et al., "A systematic review: Virtual-reality- based techniques for human exercises and health improvement," Front. Public Heal., vol. 11, 2023, doi: 10.3389/fpubh.2023.114394.

G. A. Noghre, A. D. Pazho, and H. Tabkhi, "PoseWatch : A Transformer-based Architecture for Human- centric Video Anomaly Detection U**sing** Spatio-temporal Pose Tokenizati