International Journal on Smart & Sustainable Intelligent Computing
Vol. 01, Iss. 01, S. No. 005, pp. 68-75, July 2024
ISSN (E):3048-8508

From Pixels to Perfect Form: Deep Learning for Real-time Yoga Pose Analysis

Preeti Garg[1], Karnika Dwivedi[1], Bharti Chugh[1], Madhu Gautam[1]

[1]Department of Computer Science & Engineering, KIET Group of Institutions, Delhi-NCR, Ghaziabad, India
Preeti.itgarg@gmail.com

How to cite this paper: Preeti Garg, Karnika Dwivedi, Bharti Chugh, Madhu Gautam, "From Pixels to Perfect Form: Deep Learning for Real-time Yoga Pose Analysis," *International Journal on Smart & Sustainable Intelligent Computing* , Vol. No. 01, Iss. No. 01, pp. 68–75, July 2024.

Received: 28/06/2024
Revised: 20/07/2024
Accepted: 25/07/2024
Published: 31/07/2024

Abstract

For maintain a healthy life, yoga plays a very important role. Now, a days people are using online platforms to learn yoga poses and maintain the healthy life. While doing yoga the most important point to take care is maintaining the correct yoga posture. In order to learn how to recognize suitable yoga poses and give feedback to improve posture, this research presents a deep learning-based method for estimating yoga posture. The model is periodically fed frames from videos or pictures. Key points are extracted with Keras multi-person pose estimation, yielding 12 joint vectors. The angles of these vectors relative to the x-axis are calculated. One of the six yoga positions is identified by applying a classification model based on these angles. A dataset with six distinct yoga poses, resulting in 70 videos and 350 instances, is used to test the methods. There are thirty validation examples, three hundred training instances, and thirty testing instances in the dataset. Various approaches such as Support Vector Machines (SVM), Convolutional Neural Net-works (CNN), and CNN in conjunction with Long Short-Term Memory (CNN + LSTM) are implemented, and results reveals a competitive performance mosaic. Though CNN and CNN + LSTM architectures are believed to be superior, the updated feature set allows Multi-Layer Perceptron (MLP) to achieve an impressive accuracy of 0.9958.

Keywords

LSTM, Multi-Layer Perceptron, Speed Up Robust Feature (SURF), Yoga Pose, Deep Learning

1. Introduction

Yoga is an advanced physical workout that has its origins in more than 5,000 years of spiritual history. It is well known to provide a host of health benefits for the body, mind, and soul. It fortifies the body and mind with postures that have evolved over the last few decades. Yoga is essential when done under the supervision of a trained instructor because improper alignment can lead to a number of health problems, such as pulled muscles, sprained ankles, stiff necks, etc [1]. This suggests that a teacher needs to be there to supervise the exercise and correct the participant's posture. Online education and self-study yoga platforms have expanded the availability of high-quality yoga instruction. Practitioners can select the most effective tactics with the aid of applications that analyse posture using camera vision. You can only become an independent yoga practitioner after receiving enough training. These days, more and more people prefer to practice yoga at home. Since not all clients approach or have access to trainers, a computerized reasoning-based application could identify yoga poses and offer personalized guidance to help people improve their structure [2].

Mobile devices and virtual coaching software can be utilized for this. The proposed AI-powered application can recognize different yoga poses and give users tailored advice on how to correct their posture. With applications in behaviour analysis, intelligent driver assistance systems, assisted living, and visual surveillance, human posture estimation is a hot topic in computer vision. With the advent of deep neural networks, pose estimation has advanced dramatically. Yoga poses can be standardized and corrected with the use of computer vision. As per the Narayana, performing incorrect yoga poses can result in serious injuries and persistent issues. Human posture analysis can help with at-home health by identifying and repositioning posture abnormalities. A posture estimator is used to extract yoga asana properties from precisely rendered photos.

Neural networks and ML models are immediately fed these gathered characteristics. These models evaluate and forecast the accuracy of yoga poses. There are various potential applications of machine learning in the yoga context. From pictures or videos, ML algorithms can be trained to identify and [3] categorize various yoga poses. Applications that offer real-time feedback on pose accuracy during yoga practice can be developed using this capability. Performance tracking, Wearable technology and sensors can provide data to machine learning algorithms, which can then be used to track a variety of yoga practice metrics, including heart rate, breathing patterns, and body movements. Practitioners can use this data to monitor their development, pinpoint areas for growth, and avoid injuries [4]. ML models are able to examine the preferences, past performance, and objectives of practitioners in order to generate customized suggestions for yoga asanas, poses, or adjustments based on their unique requirements. With the use of machine learning, yoga practitioners can be guided through sessions by virtual instructors who provide real-time feedback, adjustments, and modifications based on their performance A variety of large datasets pertaining to yoga, including research papers, practitioner demographics, and health outcomes, can be analyzed through the application of machine learning techniques. This analysis can provide light on the efficacy of various yoga techniques, their effects on mental and physical well-being, and the variables affecting practice adherence. All things considered, machine learning has a great deal of potential to improve the efficacy, accessibility, and customization of yoga practice, which will lead to its widespread acceptance and beneficial effects on people's health.

The objective of this research to detect the correct postures by calculating the angles so that proper training can be provided to the users. To achieve these goals in this work SVM, CNN model and CNN model along with LSTM have been implemented to detect the correction of the yoga posture and their results are compared in terms of the accuracy value. The paper is organized in various sections. Section 2 provides the literature review, the methodology is given in Section 3. Section 4 describes the result section and Conclusion is given in Section 5.

2. Related Work

This section provides the summary of various works done in this field. The study [5] proposes a deep learning-based method for detecting and correcting these poses that achieves an impressive 0.9958 accuracy with low computational complexity. To have their videos evaluated and get advice on how to straighten their posture, users can upload videos of themselves to the system. With the use of deep learning and key point detection techniques, the technology offers a solution to the issues of busy schedules and instructor accessibility limitations. The methodology includes pose classification using MLP models, joint vector computation, and point extraction. Real-world runtime analysis is used in the study to demonstrate the method's effectiveness. All things considered, the study offers a workable technique for yoga practice, with potential applications in activity recognition and sports monitoring.

The study [6] finds that MediaPipe is the most accurate architecture, particularly when using a single camera setup, after analysing four deep learning architectures on 6,000 images of yoga poses. It discusses the challenges of modelling the human body as well as the advantages of deep learning for pose estimation. The results show just how much better Media Pipe is. This study also reviews the literature on 3D human pose estimation and its applications to fitness and health. It also provides a comparative analysis of yoga interventions and an orientation program on the eight limbs of Ashtanga yoga. The value of accurately estimating yoga poses with machine learning is emphasized in the study's conclusion, and MediaPipe is the suggested architecture.

In order to enhance yoga pose classification, the study [7] presents LGDeep, a model that combines feature extraction (LDA, GDA) and deep learning (Xception, VGGNet, SqueezeNet) techniques. It performs better than other methods and demonstrates noticeable improvements in accuracy on a publicly available dataset. The study suggests computerized systems for posture recognition and personalized guidance, and it highlights the importance of standard yoga poses for pre- venting injuries and improving overall wellbeing. The method combines transfer learning with individual model training as well as ensemble training. LGDeep offers practitioners improved health and safety and outperforms previous models with 100% accuracy. The study compares favourably with previous methods and goes beyond simply identifying yoga poses.

Authors in [8] provides a half-breed deep learning demonstration for CNN and LSTM combined real-time yoga recognition from recordings. Post-polling exactness rates of 99.38% ad- dress the unoccupied space of pose acknowledgment in yoga, surpassing the single-frame precision of 99.04%. The success of the show is demonstrated by its accuracy of 98.92% in real-time testing with a wide range of subjects. The study uses sports and rehabilitation research to highlight the possible applications of computer-assisted self-training systems. The dataset is now accessible through the public domain, enhancing reproducibility and enabling greater reflection on action acknowledgment. All things considered, the ponder presents a novel strategy for identifying yoga postures and illustrates the adequacy of cutting-edge technologies in this field.

This systematic review [9] examines the use of Human Pose Estimation (HPE) in Sport and Physical Exercise (SPE) by re- viewing the literature, methods, performance, and challenges. Twenty papers were chosen from over 500 that had been sifted through databases such as dblp, Web of Science, and ACM Digital Library. While some of the data is already publicly available, more is required for robust performance across contexts. Writers frequently use Openpose or other general- purpose systems customized for particular tasks. The review ends with a discussion of the opportunities, difficulties, and constraints that HPE presents for SPE, underscoring the field's increasing significance and promise.

The authors in [10] offers a hybrid deep learning model that recognizes yoga poses in real time from recordings by combining CNN and LSTM. Post-polling precision rates of 99.38% address the modern space of pose acknowledgment in yoga, outperforming the single-frame precision of 99.04%. The model's viability is demonstrated by its 98.92% exactness in real-time testing with a variety of subjects. The conversation looks at studies on sports and therapy to show how computer- assisted self-training systems could be applied. Reproducibility has increased and more research is being done on movement recognition since the dataset is now publicly available. All things considered, the question appears to respect recent advances in this field and offers a novel approach to distinguishing between different yoga poses.

The usefulness of yoga and mindfulness as adjunctive therapies for severe mental illnesses (SMIs), such as major depressive disorder (MDD), schizophrenia, and bipolar disorder (BD), is examined in [11]. It contains studies from the past ten years that show improvements in cognition, functioning, and symptoms. Though more research is needed in these areas, yoga and mindfulness have been shown to be beneficial in improving functioning in schizophrenia and lowering depressive symptoms in MDD. The evidence for BD is weak. The review provides insight into the potential of these interventions in the treatment of SMI through discussion of participant characteristics, intervention techniques, and impacts on various variables. It also makes recommendations for future lines of inquiry and medical applications.

In the literature it has been found that DL models have been implemented in the area but there are some shortcomings in terms of limited dataset and accuracy. The proposed work has been implemented on video dataset to correctly identify the proper poses of Yoga. The work has also been implemented in real-time by capturing the video of a person and then detecting the correct pose. In the research various deep learning approaches such as SVM, CNN, CNN+LSTM and MLP have been implemented along-with their works are compared to show which approach gives better result in terms of the accuracy.

3. Methodology

3.1. Dataset Acquisition and Preparation

Five distinct yoga poses—downward dog pose, goddess pose, tree pose, plank pose and the warrior pose[19]-were included in the dataset that is used for this study. Before passing data to the model the dataset is checked for missing value, junk value and balancing. The dataset is free from missing value and junk values and the dataset is imbalanced, so dataset is balanced using SMOTE to overcome the issue of overfitting and split ratio is 70-30.The 70 videos that comprised this dataset included 350 instances of these poses. The subjects were recorded at a distance of 4 meters [11-12] using a standard RGB webcam operating at a frame rate of 30 frames per second [5]. Participants executed the poses with only minor modifications to ensure robustness in trained models. The dataset was then divided into 30 instances for validation and 320 instances for training. Additionally, a separate testing dataset comprising thirty examples—five for each yoga pose—was laboriously chosen at various intervals from the videos.

3.2. Pose Estimation

To locate and obtain the coordinates of important body points, pose estimation algorithms are employed. For real-time pose estimation, Keras multi- person pose estimation technique are used. Every two seconds, a frame was taken from the video using this method. As a result,350 instances in all were extracted from 70 videos; each video included a set of 18 critical points, each of which had an associated (x, y) coordinate. The Keras pose estimation technique is improved to more reliably rank critical points in order to improve accuracy.

3.3. Feature Extraction

Feature extraction involved the derivation of angles [13-14] between body joints using the (x, y) coordinates of key points. To obtain the vectors representing the joints, the vectors corresponding to joint positions are subtracted. For example, if the coordinates of the right shoulder and neck were denoted as (x1, y1) and (x2, y2) respectively, their corresponding vectors would be x1i + y1j for the neck and x2i + y2j for the right shoulder. Subsequently, we obtained the vector representing the joint connecting the neck and right shoulder by subtracting the neck vector from the shoulder vector [15-16] are obtained using eq 1.

$$Joint\ vector = Shoulder\ vector - Neck\ Vector$$
$$(x2 - x1)i + (y2 - y1)(-1)j \qquad (1)$$

Here, the (-1) is used to find the amount of difference in the position of the origin between the images.
Using this method 12 angles are derived for the 12 different vectors representing the 12 joints, resulting in a feature set with 12 columns.

3.4. Pose Classification

To calculate the mean values and angles of the extracted features for every pose in the dataset, the data from each individual is combined. These retrieved features are used to train a classification model to classify input images of yoga poses. In order to determine differences for each angle, the model extracted angles from the input images and compared them with precomputed average values during the inference process. The extent to which each individual needed to modify their pose was indicated by the size of these variations, which functioned as markers of pose deviation. In addition, the direction of the deviation—which was determined by the sign of the difference values—provided practitioners with personalized guidance when it came to recommendations for joint adjustments.

4. Result Analysis

An essential part of neural network architecture, the Multi- layer Perceptron (MLP) is made up of linked layers of neurons designed to handle complicated data. The network needs the input, hidden, and output layers among other crucial parts in order to recognize intricate patterns in the data landscape. The configuration and presence of hidden layers have a big influence on how well the model comprehends the data's underlying structure. A model that has too many hidden layers may overfit, forgetting the training set at the expense of applying the model's learning to new examples, while a model with too few hidden layers may underfit, failing to capture crucial details.
This work uses angular computations between important anatomical points in an attempt to leverage the power of MLP in the field of human pose classification. The input data contains a total of twelve features, six different pose classes defining different body configurations. This suggests that for every pose class, an output layer dimensionality of six is required. Two hidden layers, each with 10 and 8 neurons, reinforce the architecture. Then, in accordance with the current classification task, an output layer consisting of six neurons is added. A carefully selected 350-item dataset is used to support the model's learning process, with 30 samples going through rigorous validation and 320 samples reserved for training as shown in Table 1.

Table 1: Data Detail for Yoga Pose analysis

No. of Features	12
No. of Poses	6
No. of Hidden Layer with 10 neurons	1
No. of Hidden Layer with 8 neurons	1

No. of Output layer with 6 neurons	1
No. of Epochs	10000
Loss Function	Categorical cross-entropy
Activation Function	SoftMax
Optimizer	AdaDelta
No. of Training samples	30
No. of Validation samples	320

Mini-batches of twenty instances each can be used for iterative refinement over the 10,000 epochs that comprise the path to model mastery. After some initial fluctuations, the training and validation accuracies stabilize around epoch 6900 and reach an impressive peak of 0.9958. This stabilization is reflected in the loss function's steady decline, which shows how well the model has converged to an ideal solution.

Categorical cross-entropy is the best loss function for the multi-class classification paradigm used in the classification of human poses. The adaptive feature of the AdaDelta optimizer allows it to adjust to the evolving training process, which aids in navigating the intricate web of learning rate decay and selection. The softmax activation function in the output layer [17] is essential because it allows the model to transform raw inputs into probabilistic confidence scores for every pose class and make decisions based on the highest level of confidence. Comparison with other approaches, such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and CNN in conjunction with Long Short-Term Memory (CNN + LSTM) [18], reveals a competitive performance mosaic. Though CNN and CNN + LSTM architectures are believed to be superior, the updated feature set allows MLP to achieve an impressive accuracy of 0.9958 as shown in Table 2, matching or surpassing its more well-known competitors.

The evaluation process is viewed using the confusion matrix [16], a trustworthy tool that enables a detailed analysis of classification effectiveness. Each cell in a 6 × 6 matrix represents the model's discriminative power and classification accuracy for a given pose class. The confusion matrix is shown in Figure 1 for training, validation and testing sets. The neck and right shoulder vectors represented as coordinate vectors are shown in Figure 2 (a) and (b). The proposed model has achieved high accuracy with very less loss as shown in Figure 3 and 4.

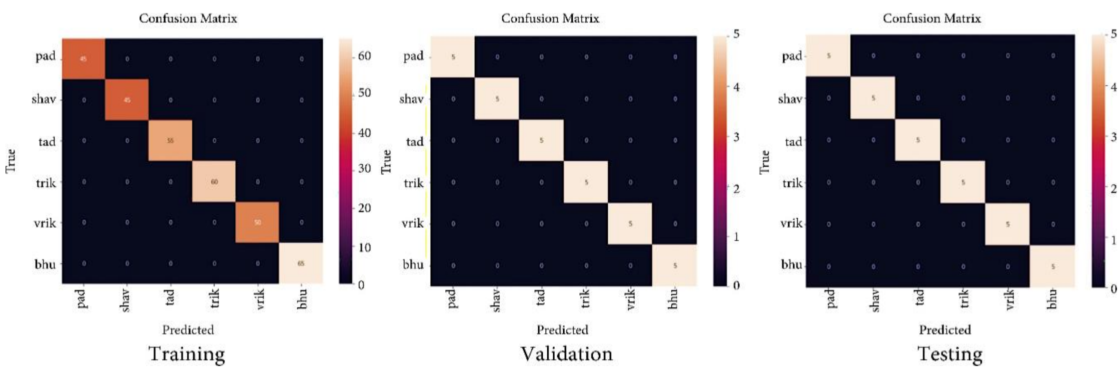
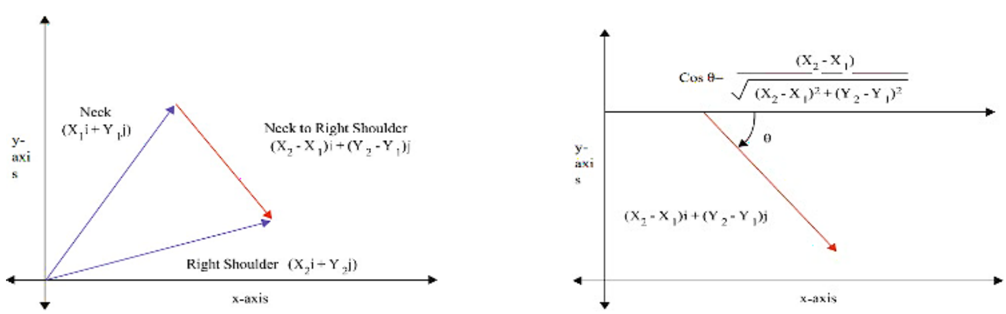


Figure 1. Confusion matrices of training, validation, and testing datasets



(a) (b)

Figure 2. The neck and right shoulder vectors represented as coordinate vectors shown in (a) and angles calculated in (b)

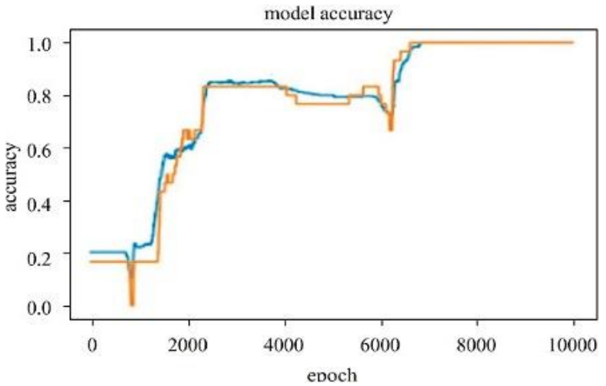


Figure 3. Accuracy graph for training and testing datasets

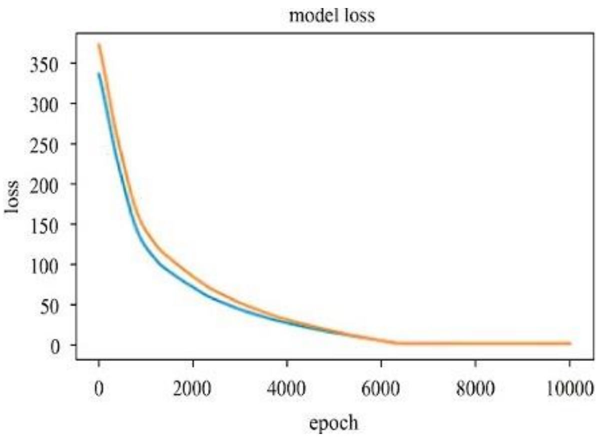


Figure 4. Loss graph for training and testing datasets

Table 2: The accuracy results of the tested models

Model	Training	Testing
SVM	0.9532	0.9319
CNN	0.9934	0.9858
CNN +LSTM	0.9987	0.9938
MLP	0.9962	0.9958

5. Conclusion

This work employs deep learning techniques to detect incorrect yoga poses and identify specific issues, as well as provide recommendations for remediation. Users can record their yoga sessions with videos and select which poses to practice. The system analyses angles as features obtained from activity monitoring, but because important points rotate, angle measurements are inaccurate. However, because this method ignores angles between joints and only accounts for angles with respect to the ground, small rotations of significant points may affect accuracy. Experiments on datasets show that a multilayer perceptron (MLP) trained with these features achieves an astounding accuracy of 99.58%. Overall, the findings of the proposed method contributing on getting better pose estimation results in comparison to other existing methods. The convolution neural networks are enough capable for extracting the meaningful feature. On top of that the integration of LSTM with CNN model helps for extracting the more relevant features from the considered dataset and MLP have proven the best the results which can help many people in maintaining a happy and healthy life. For the real-time implementation practitioners can expand this model by incorporating with smart wearable devices to make it more generalize which can help them to adjust their pose for better alignment or pose correction. In future work can be enhanced by incorporating personalized recommendations and real-time feedback mechanisms based on user performance and preferences may raise user effectiveness.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Aritz Badiola-Bengoa and Amaia Mendez-Zorrilla. A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise. Sensors, 21(18):5996, 2021.

[2] Varsha Bhole and Arun Kumar. A transfer learning-based approach to predict the shelf life of fruit. Inteligencia Artificial, 24(67):102–120, 2021.

[3] David Chang. Natural Language Processing and Graph Representation Learning for Clinical Data. PhD thesis, Yale University, 2021.

[4] Raviteja R Guddeti, Geetanjali Dang, Mark A Williams, and Venkata Mahesh Alla. Role of yoga in cardiac disease and rehabilitation. Journal of cardiopulmonary rehabilitation and prevention, 39(3):146– 152, 2019.

[5] Shrajal Jain, Aditya Rustagi, Sumeet Saurav, Ravi Saini, and Sanjay Singh. Three-dimensional cnn-inspired deep learning architecture for yoga pose recognition in the real-world environment. Neural Computing and Applications, 33:6427–6441, 2021.

[6] D Mohan Kishore, S Bindu, and Nandi Krishnamurthy Manjunath. Estimation of yoga postures using machine learning techniques. International Journal of Yoga, 15(2):137–143, 2022.

[7] Yinsong Liu, Junsheng Yu, and Hanlin Mou. Photoplethysmography- based cuffless blood pressure estimation: an image encoding and fusion approach. Physiological Measurement, 44(12):125004, 2023.

[8] Zhiqiang Luo, Weiting Yang, Zhong Qiang Ding, Lili Liu, I-Ming Chen, Song Huat Yeo, Keck Voon Ling, and Henry Been-Lirn Duh. “left arm up!” interactive yoga training in virtual environment. In 2011 IEEE virtual reality conference, pages 261–262. IEEE, 2011.

[9] Siddharth Patil, Amey Pawar, Aditya Peshave, Aamir N Ansari, and Arundhati Navada. Yoga tutor visualization and analysis using surf algo- rithm. In 2011 IEEE control and system graduate research colloquium, pages 43–46. IEEE, 2011.

[10] Ronald Poppe. A survey on vision-based human action recognition.
Image and vision computing, 28(6):976–990, 2010.

[11] Fazil Rishan, Binali De Silva, Sasmini Alawathugoda, Shakeel Ni- jabdeen, Lakmal Rupasinghe, and Chethana Liyanapathirana. Infinity yoga tutor: Yoga posture detection and correction system. In 2020 5th International conference on information technology research (ICITR), pages 1–6. IEEE, 2020.

[12] Gopinath Sathyanarayanan, Ashvini Vengadavaradan, and Balaji Bharad- waj. Role of yoga and mindfulness in severe mental illnesses: A narrative review. International journal of yoga, 12(1):3–28, 2019.

[13] Sumeet Saurav, Prashant Gidde, and Sanjay Singh. Exploration of deep learning architectures for real-time yoga pose recognition. Multimedia Tools and Applications, pages 1–43, 2024.

[14] Marc B Schure, John Christopher, and Suzanne Christopher. Mind–body medicine and the art of self-care: teaching mind-fulness to counseling students through yoga, meditation, and qigong. Journal of Counseling & Development, 86(1):47–56, 2008.
[15] Mayank Singh, Pawan K Gupta, Vipin Tyagi, Jan Flusser, Tuncer O¨ ren, and Rekha Kashyap. Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019, Ghaziabad, India, April 12– 13, 2019, Revised Selected Papers, Part II, volume 1046. Springer, 2019.
[16] Vivek Anand Thoutam, Anugrah Srivastava, Tapas Badal, Vipul Kumar Mishra, GR Sinha, Aditi Sakalle, Harshit Bhard-waj, and Manish Raj. Yoga pose estimation and feedback generation using deep learning. Computational Intelligence and Neu-roscience, 2022, 2022.
[17] Aman Upadhyay, Niha Kamal Basha, and Balasundaram Ananthakr- ishnan. Deep learning-based yoga posture recognition using the y pn- mssd model for yoga practitioners. In Healthcare, volume 11, page 609. MDPI, 2023.
[18] Santosh Kumar Yadav, Amitojdeep Singh, Abhishek Gupta, and Jagdish Lal Raheja. Real-time yoga recognition using deep learning. Neural computing and applications, 31:9349–9361, 2019.
[19] https://www.kaggle.com/datasets/niharika41298/yoga-poses-dataset?resource=download