A Novel Approach to Classify Yoga Poses by Employing Convolution Neural Network

Megha Rani Raigonda
Assistant Professor, Department of CSE (MCA), Visvesvaraya Technological University, Centre of PG Studies, Kalaburagi, Karnataka, India

ORCID iD: https://orcid.org/0000-0001-9964-3265

Rajkumar Bainoor
Assistant Professor, Department of Electronics and Communication Engineering, PDA College of Engineering, Kalaburagi, Karnataka, India

Abstract: Yoga is an oldest Indian discipline that has gained popularity throughout the world as a way to improve mental, spiritual and physical well-being. However, doing yoga incorrectly can cause numerous health related issues such as back pain, muscle strains, joint injuries, and neck or shoulder problems. Due to which, it is important to have a trainer to guide and check the correct poses of yoga. To ensure that the proposed model was designed to identify different yoga postures accurately. So, in this work 11,394 images with 10 different yoga poses were considered for classification namely ardha chandrasana, downdog, goddess, marjaryasana, padmasana, plank, tree, urdhva mukha svanasana, utkatasana and warrior 2 and employ the Convolutional Neural Network (CNN) with 11 layers and trained the model for 30 epochs. The designed system of yoga pose recognition system achieved overall accuracy of 97.14% and calculated the performance using various measures like accuracy, F1-score, precision, and recall to ensure robustness and reliability of the system.

Index Terms: Yoga Pose Recognition, Convolutional Neural Networks, Deep learning and Image processing,.

1. Introduction

Yoga is a practice that has started thousands of years ago in India. It cleanses the mind, soul and body and helps to improve physical health. It can heal various health related issues without the use of medicines. But today people adopt a sedentary lifestyle due to many factors, such as many jobs requiring long working hours, especially in office environments, increased use of computers, smartphones and other digital devices for work and entertainment, people living in cities are less active and are more dependent on cars and public transport. It plays a main role in reducing the outcomes of a sedentary lifestyle. Regular practice of yoga increases flexibility and range of motion, which reduces the stiffness associated with prolonged sitting. It emphasizes alignment and posture, helping to correct imbalances caused by prolonged sitting.

In the latest years, human posture classification has benefited greatly from deep learning and huge gains in performance have been achieved. Deep learning approaches provide a more straightforward way of mapping the structure instead of having to deal with the dependencies between structures manually [1]. Yoga became popular during the pandemic because to its convenience, ease, and affordability. This led to the develop effective yoga automation, including pose identification using machine learning.[2]. In computer vision recognition of yoga pose is a important topic having various applications in intelligent driver support systems, assisted living, to figure out the behavior, and visible surveillance [3]. The evolution in posture detection is due to the increase in the computational capability of modern computers and recent development in deep learning [4]. Image processing is essential in yoga pose recognition as it enhances image quality, extracts key body landmarks, and facilitates precise pose estimation. By preprocessing images and detecting crucial features like joints and limbs, it improves the pose recognition accuracy. Image processing and deep learning are crucial for yoga pose recognition by enabling accurate and automated recognition of complex poses. CNN model learns patterns in images for precise classification, while image processing enhances data quality and detects key points of body. Together, they facilitate reliable and real time detection of pose in fitness applications.

1.1 Outline

The remaining contents of the paper are structured as follows. In Section. 2 the literature review and contribution to knowledge is discussed. Section 3 provides the methodology, which involves the details of the dataset, preprocessing, segmentation, extraction of features and CNN used for pose recognition of yoga. In Section 4, the model training and performance evaluation results, model optimization details are presented. The concluding part of the work is explained in section 5.

2. Literature Review

Yoga Pose Recognition is a new domain in the field of deep learning. Researchers devoted their effort and time for identifying and overcoming the obstacles to classify poses of yoga. This literature review shows the works done in yoga posture classification and the challenges that they faced in classifying the posture of the human body. Deepak k, Anurag Sinha [5] suggested a method for classifying yoga in which they collected the dataset using a webcam. The CNN and LSTM applied for training the model with 6 type of yoga pose. Sumeet Saurav et al [6] presented a exploration of DL architectures for real time recognition of yoga pose. They used dataset that contain 6 different poses and applied CNN, LSTM and 3DCNN model. They achieved 99.65 % accuracy from the model. Ratnesh P et al. [7] recommended interactive yoga posture detection system to identify 6 poses. To train the model the LSTM classifier is used to identify various postures with an accuracy of 92.34%.

Using the 3D CNN model, Shrajal et al [8] proposed 3D CNN based architecture for pose recognition. Using the 3D CNN model, ten yoga postures can be recognized with accuracy of 99.39%. Santosh K et al [9] worked on various methods. An image dataset that consist six yoga asanas were created using 15 individuals (ten males and five females), and used OpenPose library to take key points for the dataset found that CNN and LSTM classifier performed well. An accuracy they achieved from this model was 99.04% for videos and 98.92% for real time. J. Palanimeera et al [10] proposed a YAP_LSTM: yoga asana prediction using long short-term memory that showed LSTM method for asana poses classification. A dataset is collected from10 people (5 men, 5 women) performing every one of these ten asanas used for the recognition model and got 99.2% accuracy.

Nagalakshmi [11] developed object classification based on network with transfer learning that showed a CNN and LSTM technique classification and used three type of dataset for the classification model and obtained an accuracy of 0.993, 0.944 and 0.928 respectively. Rao Yingdong[12] worked on various CNNs in the recognition of asana poses based on OpenPose extraction of skeleton information and built the data sets of original images and the skeleton images, subsequently, VGG16, VGG19, MobileNet and the CNN built by it are used for classifying the two data sets, and obtained 94.29% accuracy of test sets. Shakti K et al [13] proposed asana classification based on DL and created their own dataset which consists of 10 different postures of asana. The methodology used was based on DL and then use it to classify yoga posture and obtained 76% of accuracy.

3. Methodology

In this study CNN is used to automatically identify and classify various yoga poses in images. CNNs are particularly effective at spotting spatial design and suitable for tasks such as image recognition. The data flow diagram of proposed methodology is shown in below figure 1.

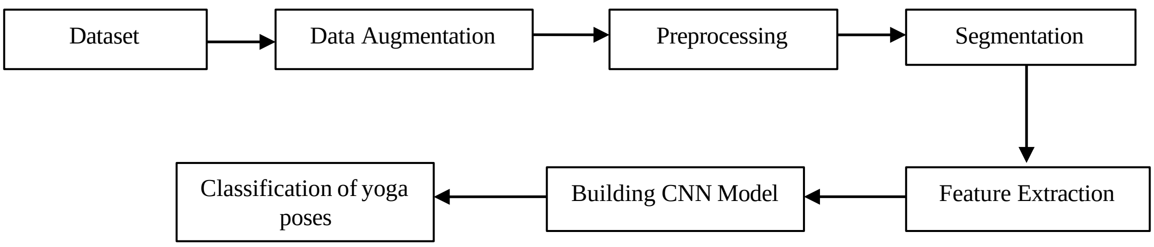

Fig. 1. Proposed Architecture of Yoga Pose Recognition System

3.1 Data Collection

The dataset used for study is a part of Open Source collection and is publicly available. It comprises images of 10 distinct yoga poses depicted in figure 2 namely ardha chandrasana, downdog, goddess, marjaryasana, padmasana, plank, tree, urdhva mukha svanasana, utkatasana and warrior2 sourced from publicly available repositories on Google and Kaggle.

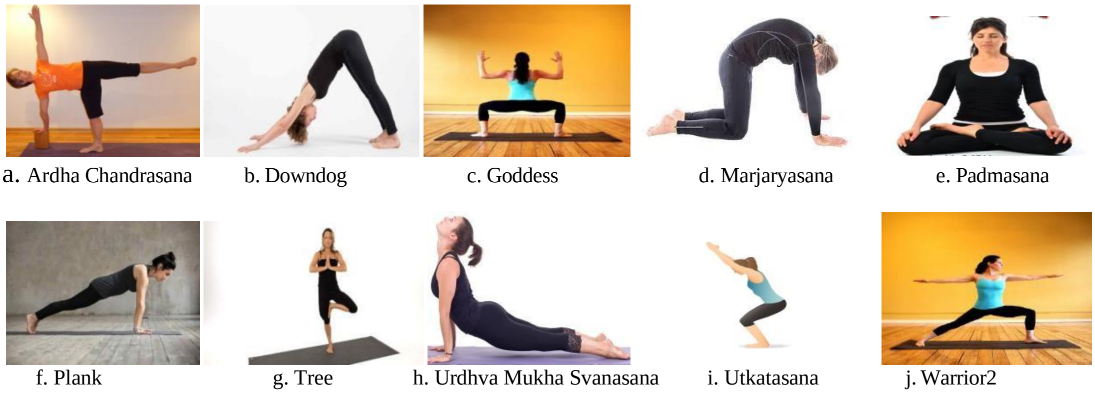

Fig. 2 Images of dataset with 10 different yoga poses.

3.2 Data Augmentation

After collecting the dataset the augmentation technique has applied on dataset by using various transformations to the original data, such as rotations, flips, zooms etc making the dataset more balanced and wide to improve model's performance. After augmentation the dataset consists of total 11,394 images.

3.3 Preprocessing

The purpose of preprocessing is to enhance the properties of the image data by reducing noise, enhancing important details, and standardizing them for analysis. In this a image was converted into a format which is suitable for further processing, such as grayscale or filtered versions and image that has preprocessed in below figure 3 shows color normalization or contrast adjustments to make relevant features stand out better.

Fig. 3. a. Original Image b. Preprocessed Image

3.4 Segmentation

Image segmentation is the process in which the images are partitioned into several segments. It simplifies the image, making it simple to analyze by isolating regions of interest. The image of segmentation shown in figure 4 focuses on isolating the subject from the background. The technique used could be edge detection or region-based methods that defines the subject distinctly.

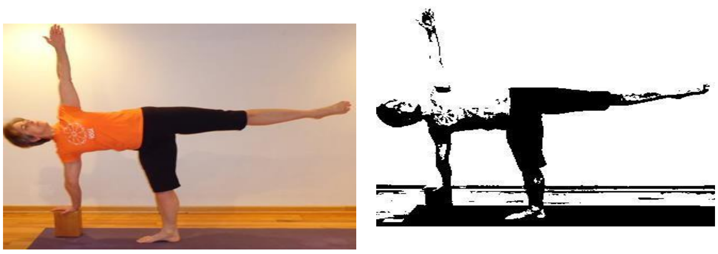

Fig. 4. a. Original Image b. Segmented Image

3.5 Feature Extraction

Feature extraction is the method of identifying specific patterns, shape, or textures from the segmented areas that can be used for further analysis. The feature extracted image in the figure 5 identifies key points, edges, or specific characteristics from the subject. This technique may involve histogram analysis, shape descriptors, or texture patterns.

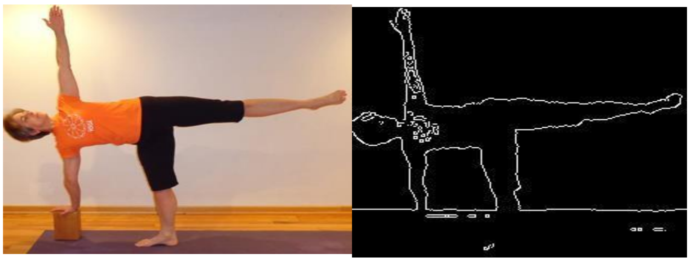

Fig. 5. a. Original Image b. Feature Extracted Image

3.6 CNN Model Development

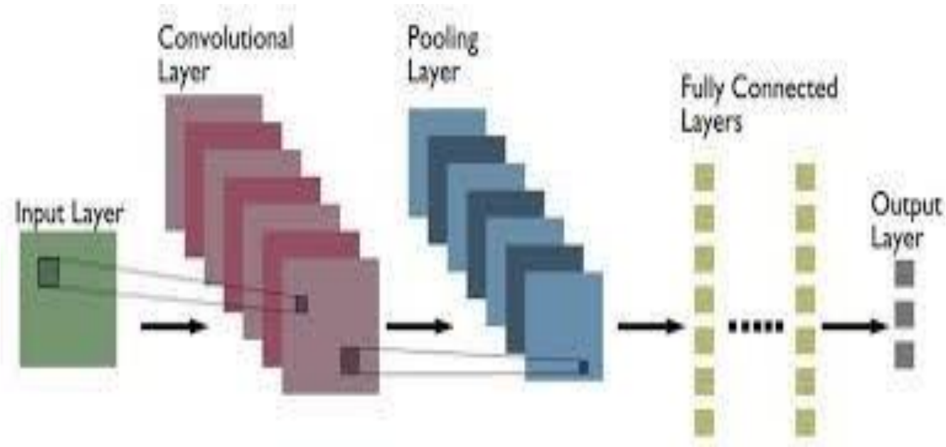

Fig. 6. Convolutional Neural Networks Architecture

A model that is proposed using CNN contains 11 layers in which first layer is 2D convolutional layer that takes the input image of size 224 x 224 with 3 color channels (RGB). It applies 32 convolutional filters, each of size 3 x 3, to the input image. This layer is used to detect low-level features such as edges, lines, and textures. The ReLU (Rectified Linear Unit) activation function in formula (1) introduces non-linearity to help the network learn complex patterns.

$$f(x) = \max(0, x) \tag{1}$$

Where f(x) indicates the result and x is the input of activation function called ReLu.

Similarly three more 2D convolutional layer with 64,128,128 filters respectively are used to capture more complex and high level features also ReLU is used again to introduce non-linearity. After each 2D convolutional layer the pooling layer was used that decreases the dimensionality of feature maps by taking the highest value from each 2 x 2 window. This decreases the dimensions of spatial, making the network more efficient and less prone to overfitting while retaining the most important features. After the final MaxPooling layer, feature maps of 2D are converted into a 1D vector. This layer serves as a bridge between convolution and fully connected layers. Next, fully connected (Dense) layer with 512 neurons processes the 1D vector produced by the flatten layer. Each neuron in this layer is connected to all neurons from the preceding layer, allowing the networks to learn complex combinations of the feature extracted by the covolutional layers. Finally the fully connected layer is used in which the total number of neurons are equal to total number of classes in the task of recognition in which activation function called softmax was also used which converts the output into a probability distribution, indicating the likelihood of the input of image with yoga pose belonging to each class. The pose of yoga that has highest probability is selected as predicted class.

4. Results and Discussion

The Model was developed by employing python libraries such as NumPy, Pandas, TensorFlow, PIL, OS, Matplotlib etc. The dataset contains images of 10 different yoga asanas. As mentioned earlier, the dataset consists of 11,394 yoga pose images belonging to ten classes.

After building the CNN model categorical-cross entropy loss function is used for compiling the final model which is appropriate for multi-class classification. This loss function helps in comparing the true class probabilities against the predicted probabilities and penalizes the model accordingly when the predictions deviate from the actual targets. Next, the Adam optimizer used in formula (2) and (3) dynamically adjusts learning rates during training, helping the model converge more quickly. The accuracy metric is used for evaluating the model's performance, where it calculates the percentage of correct predictions over total predictions. The model involves training over 30 epochs and it progresses through the epoch, indicating updates in these metrics as learning proceeds. The accuracy and loss curves are presented in figure 7 and figure 8 where we can see that the model (Convo2d) does not suffer from over-fitting. The loss curves of the proposed model and accuracy are smooth in trajectory.

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\left[\frac{\delta L}{\delta w_t}\right] \quad (2)$$

Where m_t indicates aggregate of gradients at time t, m_{t-1} indicates aggregate of gradients at time t-1, w_t indicates weight at time t, δL indicates derivatives of loss function, δw_t indicates derivatives of weights at time t, and β indicates moving average parameter (const, 0.9)

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\left[\frac{\delta L}{\delta w_t}\right]^2 \quad (3)$$

Where v_t indicates sum of square of past gradients, v_{t-1} indicates sum of square of gradients at time t-1, w_t indicates weight at time t, δL indicates derivatives of loss function, δw_t indicates derivatives of weights at time t, and β represents moving average parameter.

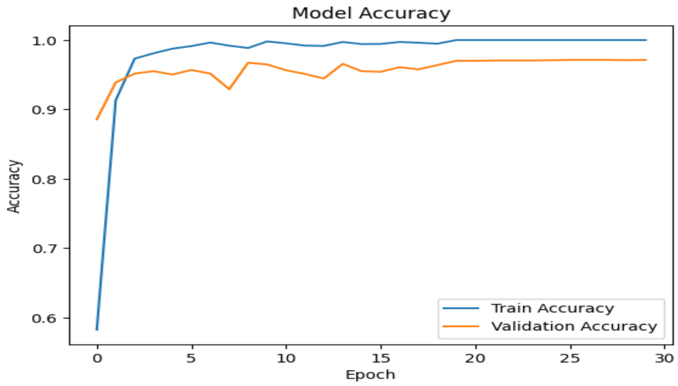

Fig. 7. Accuracy graph

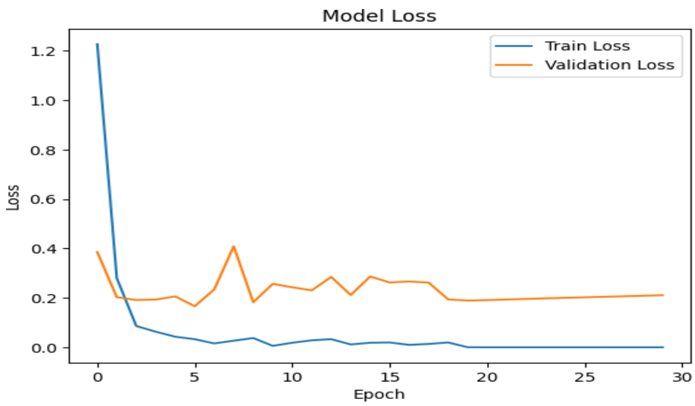

Fig. 8. Loss graph

The figure 9 presents the confusion matrices which are typically used to visualize the performance of the model by showing the counts of predicted and actual class labels of yoga pose images. Each cell represents the number of times a class was classified as another class or itself.

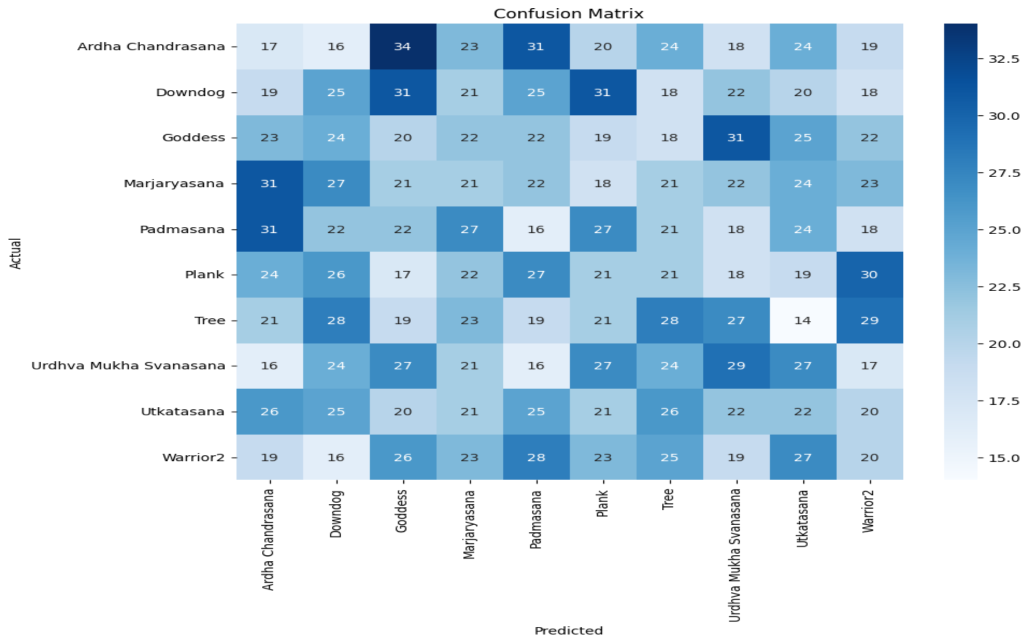

Fig. 9. Confusion Matrix

The classification report for trained model is also generated which shown in table-1 to get the analysis about how it is performing. Then using the metrics like precision, F1 score and recall, performance of classification model is evaluated. To calculate those metrics we used various formulas which are given below:

Formula to calculate precision:

The below mentioned formula (4) shows calculation of precision by dividing predictions of true positives to the summation of true positive and false positive prediction of yoga poses.

$$p = \frac{TP}{TP+FP} \quad (4)$$

Where p indicate Precision, TP indicate True Positive and FP indicate False Positive.

Formula to calculate Recall:

The formula (5) mentioned below is to compute the ratio of true positive predictions to summation of false negatives and true positives.

$$R = \frac{TP}{TP+FN} \quad (5)$$

Where R indicate Recall, TP indicate True Positive and FN indicate False Negative.

Formula to calculate F1 score:

The formula (6) is for calculating F1 Score by combining both recall and precision into a one metric.

$$F1\ Score = 2x\frac{PxR}{P+R} \quad (6)$$

Where R indicate Recall and p indicate Precision

Table 1. Classification Report

Classification Report	Precision	Recall	F1-score
Ardha Chandrasana	0.07	0.08	0.08
Downdog	0.11	0.11	0.11
Goddess	0.08	0.09	0.09
Marjaryasana	0.09	0.09	0.09
Padmasana	0.07	0.07	0.07
Plank	0.09	0.09	0.09
Tree	0.12	0.12	0.12
Urdhva Mukha Svanasana	0.13	0.13	0.13
Utkatakonasa	0.10	0.10	0.10
Warrior 2	0.09	0.09	0.09
macro avg	0.10	0.10	0.10
Weighted avg	0.10	0.10	0.10

We also compared the accuracy, F1 score, recall and precision of 2 different classifiers shown in table-2 and figure 10 to determine the most accurate classifier model on our dataset. Out of which CNN classifier showed maximum accuracy.

Table 2. Comparison of different models performance using various metrics

Algorithm	Accuracy	Recall	F1 Score	Precision
CNN	97.14	97	97	95
Decision Tree	76.86	76.86	76.87	76.96

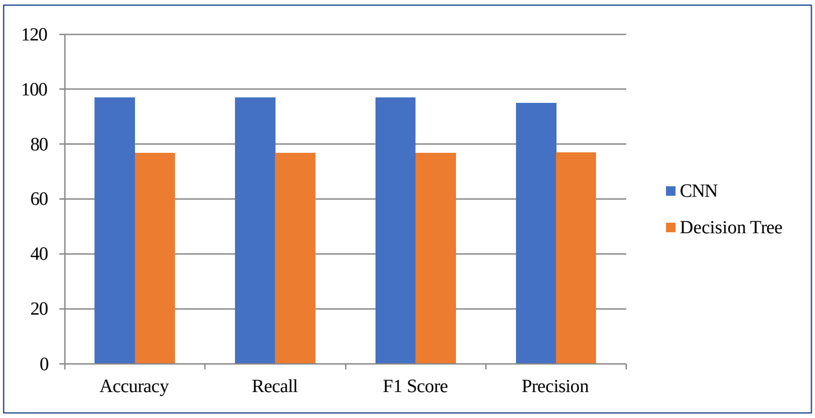

Fig. 10. Performance comparison chart of different models using various metrics

5. Conclusion

This paper shows a robust CNN-based architecture for classifying the yoga poses, demonstrating significant potential in the domain of automated fitness applications. By leveraging the deep learning capabilities of CNNs, our system achieves a high classification accuracy of 97.14% on the validation dataset, indicating its effectiveness in recognizing and distinguishing between various poses of yoga. The system proposed was not only contributes to the advancement of practicing yoga through technology but also serves as a foundation for future developments in real-time pose correction and feedback mechanisms. The results suggest that techniques of deep learning, particularly CNNs, can play a pivotal role in enhancing the precision and efficiency of fitness training systems. For Future work we can focus more on expanding the dataset to include a wide range of yoga postures and integrating real-time analysis capabilities.

References

[1] S. Kothari, Yoga poses recognition using deep learning, Master's thesis, San Jos´e State University, pp 1-45, 2020.

[2] F. B. Ashraf, M. U. Islam, M. R. Kabir, and J. Uddin, YoNet: a neural networks for yoga poses classification, SN Comput. Sci., pp 1-9, 2023.

[3] Amira S T, deep learning models of yoga pose estimator, Springer, pp. 1-13, 2023.

[4] S. Garg, A. Saxena, and R. Gupta, Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application, J Ambient Intell Human Comput, pp 16551-16562, 2022.

[5] D. Kumar and A. Sinha, Yoga Pose Detection and Classification Using Deep Learning, International Journal of Scientific Research in Information Technology and Computer Science Engineering, pp 1-25, 2020.

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

[7] Ratnesh Prasad Srivastava, Lokendra Singh Umrao and Ramjeet Singh Yadav, Real-time yoga pose classification with 3-D pose estimation model with LSTM, Multimedia Tools and Applications, pp 1-12, 2023.

[8] Sumeet Saurav, Shrajal Jain, Ravi Saini, Sanjay Singh and Aditya Rastogi, 3D CNN based Deep Learning architecture of Yoga pose recognition, Neural Computing and Applications, pp 6427-6411, 2020.

[9] Jagdish Lal Raheja Santosh Kumar Yadav, Abhishek Gupta and Amitojdeep Singh, Yoga pose recognition using deep learning in real time, Neural Computing and Applications, pp 9349-9361 May 2019.

[10] J.Palanimeera, K. Ponmozhi1, YAP_LSTM: yoga asana prediction using asana estimation and longshort-term memory, Soft Computing, pp 1-16, 2023.

[11] Nagalakshmi Vallabhaneni, Panneer Prabhavathy, Segmentation assessment based on network object detection and CNN with transfer learning for yoga pose recognition, Soft Computing, pp 1-23, 2023.

[12] Rao Yingdong, Research on Convolutional Neural Networks in Classification Scenes of Yoga Poses using OpenPose Extraction, 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), pp 1-4, 2022.

[13] Shakti Kinger, Abhishek Desai, Sarvarth Patil, Hrishikesh Sinalkar, Nachiket Deore, Using Deep Learning for Yoga Pose Classification, 2022 International Conference on Big Data, Machine Learning, Parallel and Cloud Computing (COM-IT-CON), pp 682-691, 2022.

Authors' Profiles

Dr. Megha Rani Raigonda, is working as Assistant Professor at the Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University, Centre of PG Studies, Kalaburagi, Karnataka, India. She is in teaching for more than 10 years. She has published more than 30 papers in National / International Journals and presented various papers in National / International conferences. She is author of textbook "Fundamentals of DBMS", Her main areas of interest includes Machine Learning, Image Processing, and Artificial Intelligence.

Mr. Rajkumar. P. Bainoor is Assistant Professor in the Dept. of E & CE, P D A College of Engineering, Kalaburagi, Karnataka, India. He is in teaching for more than 15 years. He has published more than 10 papers in International Conferences and journals. His main areas of interest include Power Electronics, PLC, Image Processing,Artificial Intelligence, Computer Networks, and communication System.