ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

Skeleton Pose Estimation Features based Classification of Yoga Asana using Deep Learning Techniques

S. Abarna[1], V. Rathikarani[2] , P. Dhanalakshmi[3]

[1]Research Scholar, [2]Assistant Professor, [3]Professor

[1,2,3]Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamilnadu, India.

Abstract-

Yoga is a distinct spiritual discipline of identity and personality that tells us how to maximise our potential in our many-faceted life. In recent years, yoga has become a part of living for many individuals all over the world. As a result, a scientific analysis of yoga postures is required. Pose detection techniques are useful in identifying postures and assisting people in performing yoga more precisely. Recognizing yoga pose is a difficult work due to the scarcity of datasets and the need to classify posture from a huge dataset. To address this issue, a large dataset containing six different types of yoga pose images was collected from the Yoga 82 online dataset. In the proposed work, the yoga pose is taken as input, the features were detected using pose estimation method, CenterNet and OpenPose are the pre-trained models and the deep learning algorithms namely 1D Convolutional Neural Network and Recurrent Neural Network used for classification. By comparing the experimental results the performance of OpenPose with 1D Convolutional Neural Networks yields better results when compared to other models.

Keywords- feature extraction, 1D Convolutional Neural Network, OpenPose, CenterNetand Recurrent Neural Network.

1. Introduction

Yoga's integrated method delivers profound harmony and unwavering stability to body and mind in order to activate the latent ability for higher awareness, which is the real goal of human evolution. Human position estimation and gesture identification are appealing study subjects in computer vision and robotics because of their wide range of applications, which include human-computer interaction, game control, and surveillance [5].

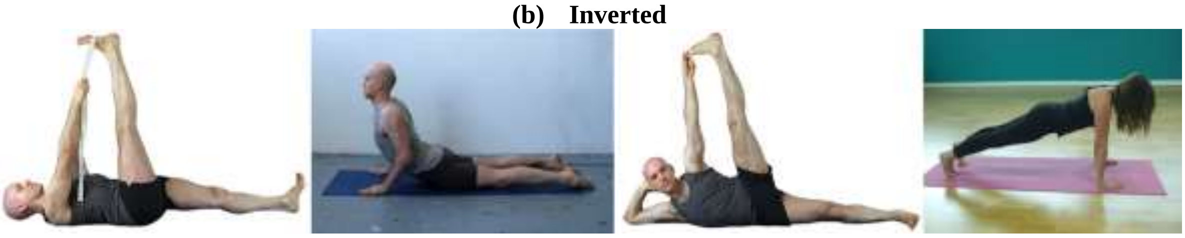
Yoga's ultimate objective is self-realization. Self-awareness is the insight that all sentient creatures are interrelated and that what we think, say, and do impacts others around us. Yoga is a fantastic kind of exercise that provides several initial and long physical advantages ranging from increased flexible to healthier muscles and bones. Yoga, on the other hand, involves more than merely going through the positions. Mindfulness is a vital component of any serious yoga practise. Yoga, when done correctly, stimulates the brain of all distracting ideas from the outer world, giving one to a position of inner serenity. [2]

1.1 Yoga Asana

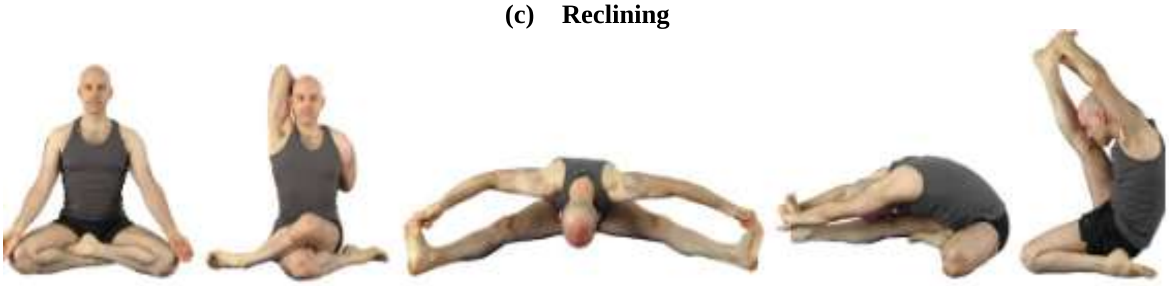
Blossoming to our fullest capabilities is the aim of Yoga. The Yoga Asanas are of two types, dynamic and static, where practice by mix. Even while the term "yoga" has become a household term, it's doesn't followed that the philosophical underpinnings of yoga are similarly generally recognised and understood. Yoga is a culture that combines wellness art, dance, music, religion, philosophy, and other areas of life. The Yoga Asanas namely Balancing, Inverted, Reclining, Sitting, Standing and Wheel type are shown in Figure 1.

(a) Balanacing

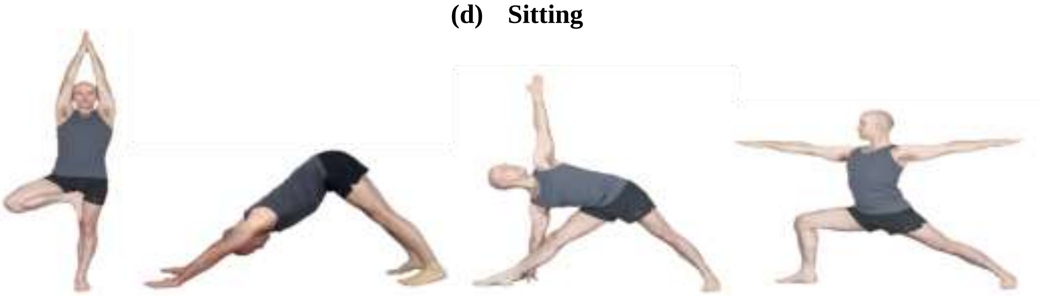
(b) Inverted



(c) Reclining



(d) Sitting



(e) Standing

(f) Wheel

In theproposed work, the pre-trained models namely, CenterNet and OpenPose are used for feature extraction and the deep learning models namely 1D Convolutional Neural Network(CNN) and Recurrent Neural Network (RNN) are used to organize features into their appropriate classes Figure 2 depicts the proposed work's block diagram.

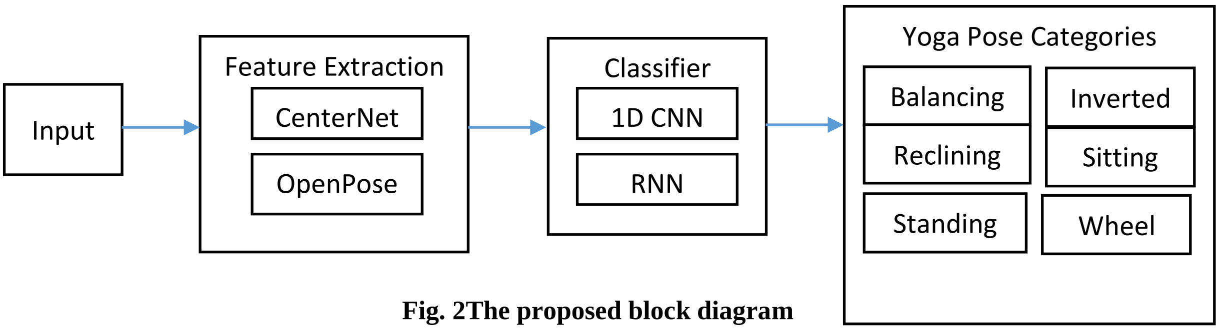


Fig. 2The proposed block diagram

2. Literature Review

The real-time dataset of 5500 images from ten various yoga postures were created with the use of tf-pose estimation algorithm. The skeletal muscle of a human body is attracted in this work, and the edges of human body joints have been extracted using the tf-pose skeleton algorithm in conjunction with various machine learning techniques such as Regression Analysis, Unusual Forest, Svm, Decision Tree, Navies Bayes, and KNN.The performance of using Random Forest Classifier gives better accuracy [3-6].

In [5] six different types of yoga asanas named asPadmasana, Bhujangasana, Tadasana, Shavasana, Trikonasana, and Vrikshasana were recognized with the hybrid approach of convolutional neural network (CNN) and long short-term memory (LSTM). The real-time video dataset is created with fifteen persons such as five femalesand ten males using a normal RGB webcam. The keypoint features of each frame are extracted in the CNN layer which is followed by LSTM for temporal predictions. This proposed work yields 99.04 % of accuracy with a single frame and 99.38 % of accuracy on 45 frames of videos.

Human interpretative process and gesture recognition techniques using support vector machine detected gestures independent of motion speed (SVM). Gesture recognition is achieved by measuring input keyframes to registered gesture keyframes. The experimental findings of this suggested study demonstrate that the strategy performs quite well and is useful in real-world settings [6-8].

The system recognizes the yoga pose from the image as well as from video by using deep learning techniques namely CNN and transfer learning. Ten different yoga asana were used to train the model and evaluate the prediction accuracy. VGG16 was used as a feature extractor which is one of the pre-trained modelsand a deep neural network classifier was used to classify the output classes. Prediction accuracy of 85% yields in this work [9-12].

3. Feature Extraction

3.1 CenterNet

CenterNet provides a fast, simple, and reliable technique for estimating the bounding boxes as well as postures of objects and people without the use of several NMS thresholds or post-processing. It treats the centre of a box both as an object and then a key point, and then utilises this anticipated centre to determine the coordinates and offsets of the bounding box.CenterNet uses a different approach than most object detection and tracking methods. Instead of enumerating through a huge list of potential object locations and sizes to classify, this model uses keypoint estimation to represent an entire object as a single point with a bounding box centered around that point. Such style affords a much higher throughput, making this model very performant.

3.2 OpenPose

Zhe Cao et al. proposed the OpenPose model in the year 2019. OpenPose is a bottom-up strategy in which the network detects body parts or key points in an image first, then maps appropriate key points to generate pairings. The main architecture of OpenPose is CNN. It is made up of a VGG-19 convolutional network that extracts patterns and representations from the input. The VGG-19's output is split into two convolutional network branches. The first network predicts a set of confidence maps for each body component, while the second predicts Part Affinity Fields (PAFs), which establishes a degree of relationship between parts. Pruning the weaker linkages in bipartite graphs is also beneficial.

4. Classifiers

In this proposedwork1DCNN and RNN are the deep learning techniques used to classify the features.

4.1 1D CNN

Convolutional Neural Network (CNN) is mainly used for audio classification, image recognition, natural language processing, medical image analysis, historical data analysis, etc. CNN is made up of three layers: the input layer, the output layer, and multiple hidden layers. Many convolution layers make up the hidden layers.

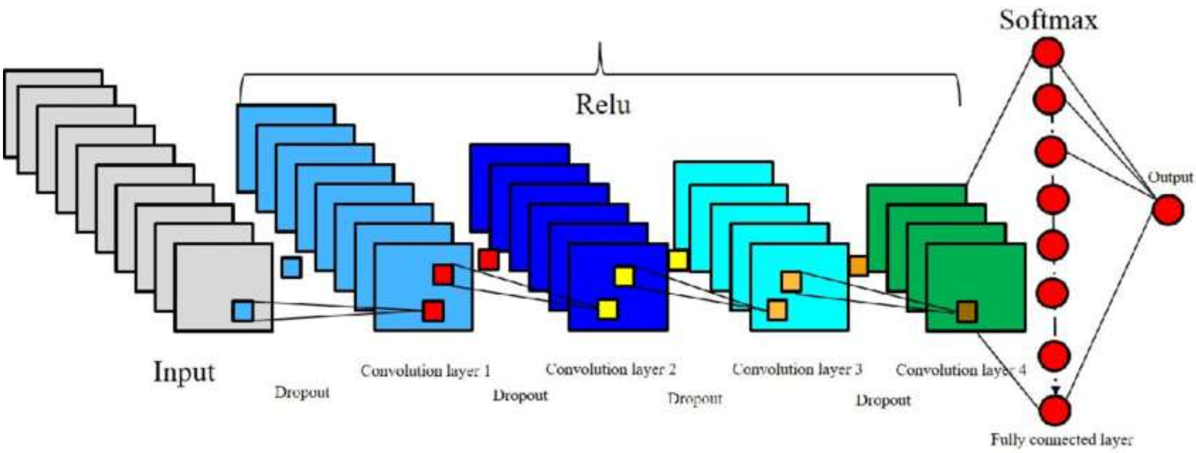


Fig. 3 CNN Architecture

The CNN model is designed with a simple stack of multiple convolutional layers followed by Rectified Linear Units (ReLUs)activation function proceeded by max pooling layers. More significantly, it is used as an optimization method. The training speed of Deep Neural Network was increased in a rapid manner usingReLUsthan applying the softmax functions [1]. The Fig. 3 shows the architecture of CNN.

4.2 Recurrent Neural Network

Recurrent Neural Networks (RNN) are still a form of neural network with hidden states and the ability to use past outputs as inputs.

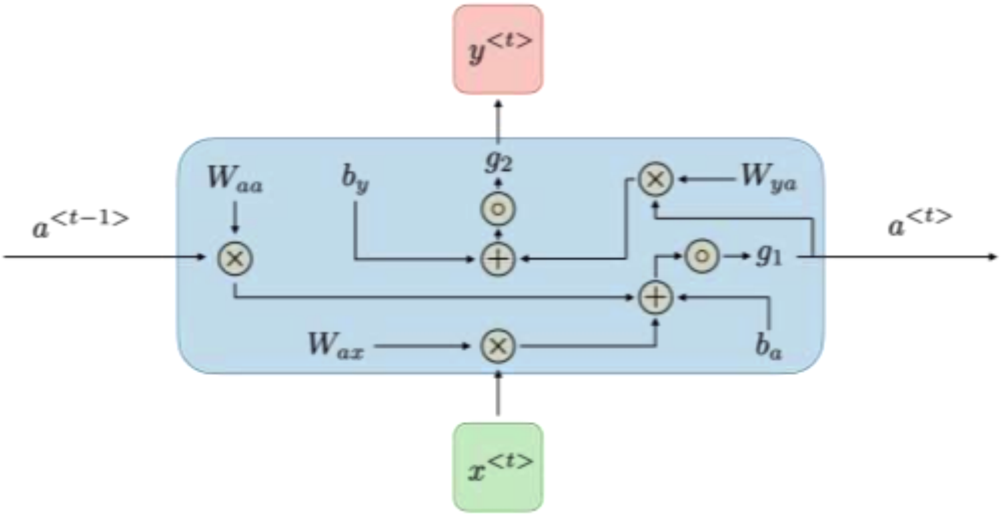


Fig. 4 Architecture of RNN model

The activation $a^{<t>}$ and the output $y^{<t>}$ are expressed as follows for each timestep *t*,:

$$a^{<t>}=g1(W_{aaa}^{<t-1>}+W_{axx}^{<t>}+b_a) \quad \text{and} \quad y^{<t>}=g2(W_{yaa}^{<t>}+b_y)$$

where W_{ax}, W_{aa}, W_{ya}, b_a, by are temporally shared coefficients and g2 activation functions**4.3 Long Short-Term Memory Units**

Long Short-Term Memory Units (LSTM) is an artificial RNN architecture that is used in deep learning fields. It solves the problem of vanishing gradients that traditional RNNs have, with LSTM being a generalization of GRU. LSTMs are especially designed to avoid the issue of long-term reliance. RNNs are now all made up of a sequence of neural network modules that are repeated. In conventional RNNs, this repeating module will have a basic structure, including a single tanh layer.Fig. 4 shows the architecture of LSTM.

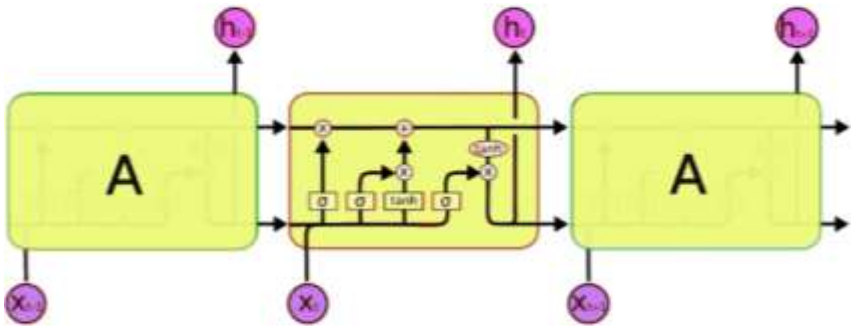


Fig. 5 Architecture of LSTM

5 PROPOSED WORK

5.2 1DConvolutional Neural Network Model

In the proposed framework, the 1 D Convolutional Neural Network (CNN)model is designed with a simple stack of convolution layers.There are 2 CNN layers are used in this work. The first convolution layer with the filter size of 512, kernel size is 1, in the next convolution layer, filter size is 256 with the kernel size is 1 and the ReLU function is activated, the output layer is the softmax function.

5.1 Recurrent Neural Network Model

In this paper,Recurrent Neural Network(RNN) structure is modeled to classify the Yoga Pose which includes the input layer, hidden layer with memory and the output layer. The proposed RNN consists of an input layer, 2 LSTM layers and an output layer. In the first LSTM layer, the output shape is 512, then the second LSTM layer the output shape is 256, next is the dense layer defines the final output shape, the dropout layer which reduces the overfitting and softmax function is the output layer.

6. Experimental Results

6.1 Datasets

The dataset was collected from Kaggle online database. The dataset consists of RGB images. 75 percent of the dataset was used for training and 25 were used for testing.

6.2 Classification using 1D CNN and RNN Model

In the proposed work, theskeleton keypoints are extracted using human pose estimation algorithm namely OpenPose and CenterNet model. Totally 34 skeleton keypoints were extracted for OpenPose and 32 skeleton keypoints were extracted for CenterNetfrom each Yoga Pose. Then the features are given into the 1D CNN and RNN model for classification. Fig. 6 shows the CenterNet features and the classification of CenterNet with the 1D CNN model shown in Fig. 7.

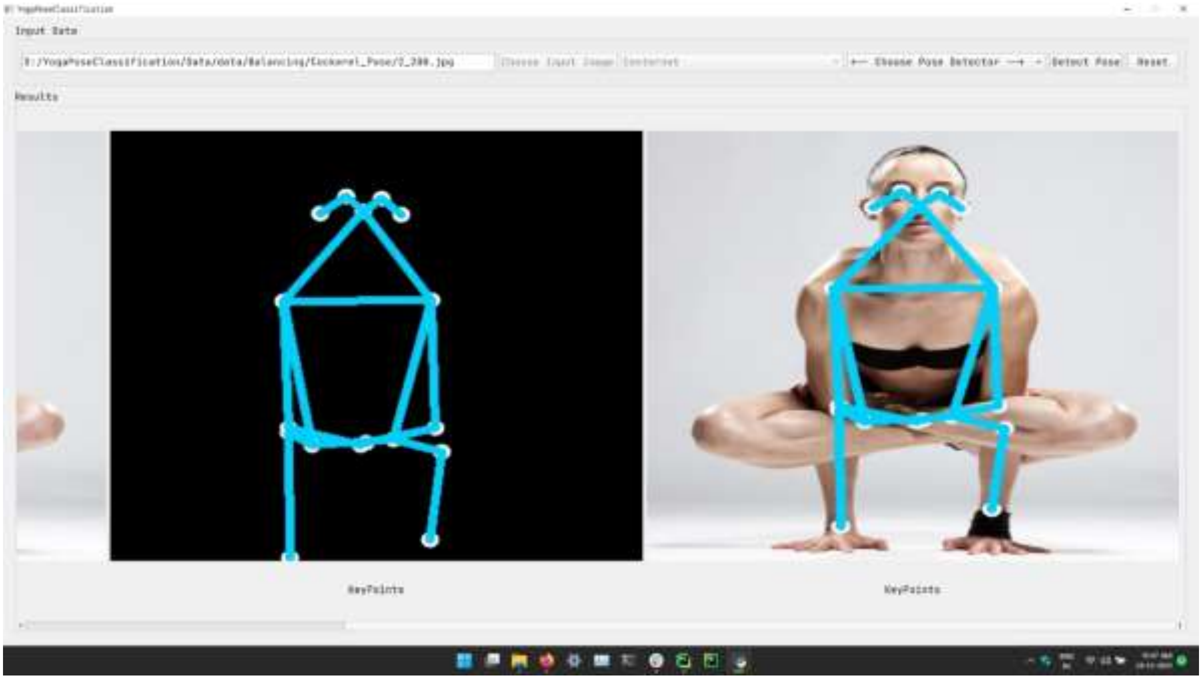


Fig. 6 CenterNet Feature extraction

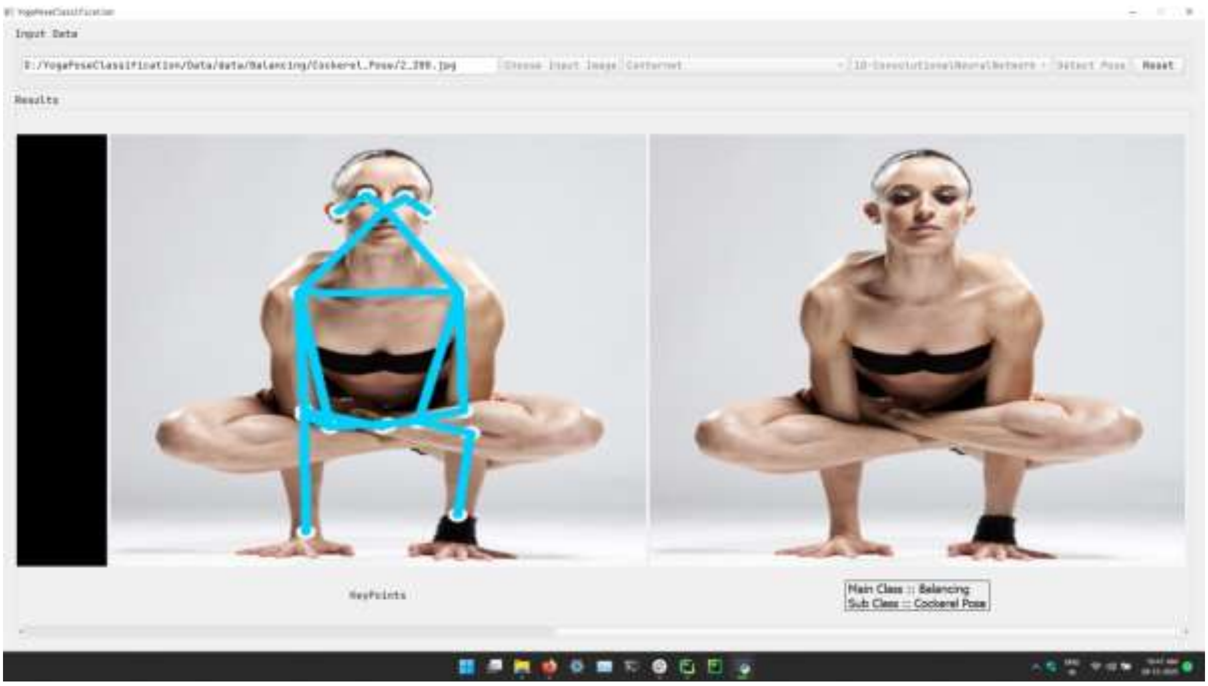


Fig. 7 Classification using CenterNet with 1D CNN

6.3Performance Measures

Accuracy, Precision, Recall, and F-score are employed as performance measurements.The overall accuracy is shown in Table 1. The accuracy of every Yoga Pose is calculated using the confusion matrix. The data can be trained and tested effectively; Precision, Recall, F-Score and Accuracy of the proposed work are

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$\text{Accuracy} = \frac{(True\ Positive + True\ Negative)}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

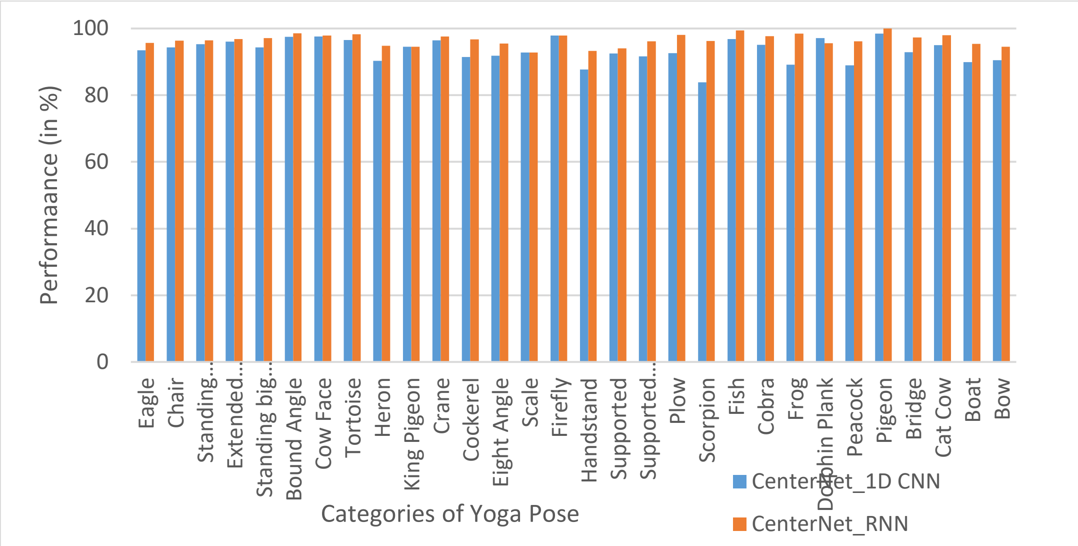


Fig. 8 Performance of CenterNet with 1D CNN and RNN

The accuracy of CenterNet and OpenPose with 1D CNN and RNN are shown in Fig. 8 and Fig. 9. The performance of OpenPose with 1D CNN gives better performance when compared to other models. The comparison of precision, recall and f-score are shown in Fig. 10. The overall accuracy of the proposed work is shown in Table 1.

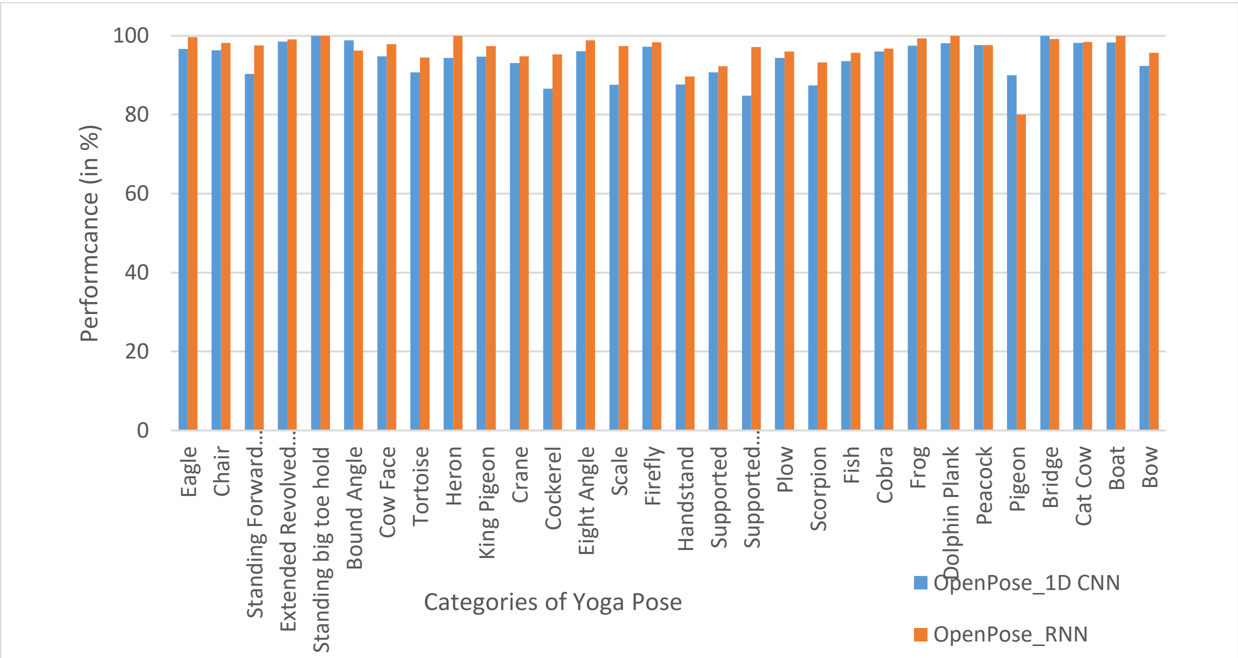


Fig. 9 Performance of OpenPose with 1D CNN and RNN

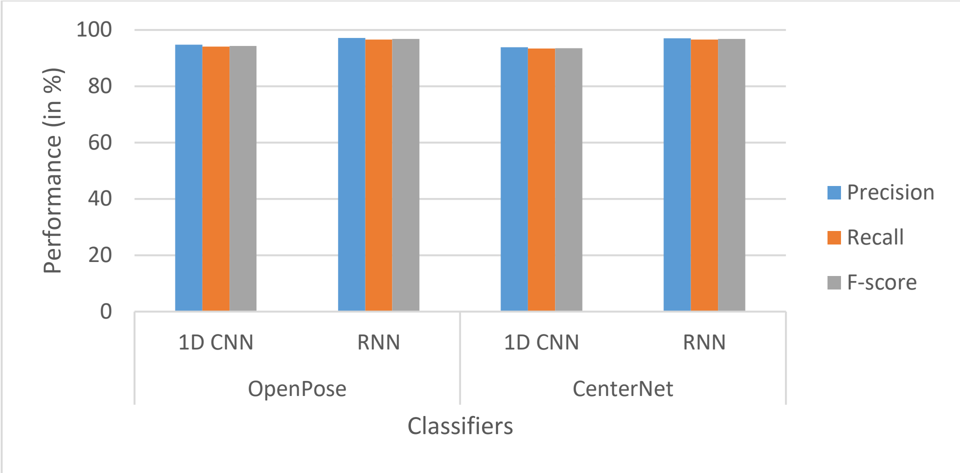


Fig. 10 Comparison of the proposed work

Table. 1 Overall Accuracy

Features	Classifiers	Accuracy (in %)
OpenPose	**1D CNN**	94.87
	RNN	86.54
CenterNet	**1D CNN**	94.23
	RNN	89.74

7. Conclusion

In the proposed work, yoga postures are classified using 1D CNN and RNN model using the CenterNet and OpenPose features. Deep learning approaches including such 1-dimensional Convolutional Neural Network & Recurrent Neural Network were proposed in this research.The skeleton keypoints were extracted from the Yoga images using the two feature extraction techniques namely CenterNet and OpenPose models which is the human pose estimation techniques. The features are classified using the deep learning techniques namely 1D CNN and RNN model. The result for the proposed work gives satisfactory results, that OpenPose with 1D CNN model gives good results when compared to the RNN model.

8. References

1. Glorot, X., Bordes, A. &Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics,pp. 315-323.
2. 2,100 Asanas: The Complete Yoga Poses, (2015), By Daniel Lacerda
3. Edwin W. Trejo and Peijiang Yuan, Recognition of Yoga Poses through an Interactive System with Kinect device, 2018 2nd International Conference on Robotics and Automation Sciences, 978-1-5386-7370-6/18, 2018 IEEE
4. Yoga for Your Type: An Ayurvedic Approach to Your Asana Practice, By David Frawley, Sandra Summerfield Kozak, 2001.
5. Asanas: 608 Yoga Poses, By Dharma Mittra, 2003.
6. Y. Agrawal, Y. Shah and A. Sharma, "Implementation of Machine Learning Technique for Identification of Yoga Poses," 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), 2020, pp. 40-43, doi: 10.1109/CSNT48778.2020.9115758.
7. Yadav, S.K., Singh, A., Gupta, A. et al. Real-time Yoga recognition using deep learning. Neural Comput&Applic 31, 9349–9361 (2019)
8. Hanguen Kim, Sangwon Lee, Dongsung Lee, Soonmin Choi, JinsunJu and Hyun Myung, Real-Time Human Pose Estimation and Gesture Recognition from Depth Images Using Superpixels and SVM Classifier, Sensors, 2015, 15, 12410-12427.
9. ExNET: Deep Neural Network for Exercise Pose Detection
10. Lowe, D. G., (2004), Distinctive image features from scale invariant key points, International Journal of Computer Vision, Vol. 60(2), pp. 91-110.
11. Josvin Jose, Shailesh S., (2021), Yoga Asana Identification: A Deep Learning Approach, IOP Conf. Series: Materials Science and Engineering 1110 (2021) 012002.
12. Islam, Muhammad Usama; Mahmud, Hasan; Ashraf, Faisal Bin; Hossain, Iqbal; Hasan, Md. Kamrul (2017). [IEEE 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) - Dhaka, Bangladesh (2017.12.21-2017.12.23)] 2017

IEEE Region 10 Humanitarian Technology Conference (R10-HTC) - Yoga posture recognition by detecting human joint points in real time using microsoftkinect. Pp. 668–673.

13. Latchoumi, T. P., & Parthiban, L. (2022). Quasi oppositional dragonfly algorithm for load balancing in cloud computing environment. Wireless Personal Communications, 122(3), 2639-2656.
14. Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2021). Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model. Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science), 14(6), 1728-1741.
15. Garikapati, P., Balamurugan, K., Latchoumi, T. P., & Malkapuram, R. (2021). A Cluster-Profile Comparative Study on Machining AlSi7/63% of SiC Hybrid Composite Using Agglomerative Hierarchical Clustering and K-Means. Silicon, 13(4), 961-972.
16. Karnan, B., Kuppusamy, A., Latchoumi, T. P., Banerjee, A., Sinha, A., Biswas, A., & Subramanian, A. K. (2022). Multi-response Optimization of Turning Parameters for Cryogenically Treated and Tempered WC–Co Inserts. Journal of The Institution of Engineers (India): Series D, 1-12.