Original Article

Classification Of Yoga Pose Using Pretrained Convolutional Neural Networks

V. Rathikarani[1]*, S. Abarna[2], K. Vijayakumar[3]

[1]*Assistant Professor, Department of Computer Science and Engineering,
[2]Research Scholar, Department of Computer Science and Engineering,
[3]Professor, Department of Library and Information Science, Annamalai University, Chidambaram, Tamilnadu, India.

***Corresponding Author: -** V. Rathikarani

*Assistant Professor, Department of Computer Science and Engineering
DOI: 10.47750/pnr.2022.13.S08.474

Abstract

Yoga is an extraordinary spiritual science of self-development and self-realization that shows us how to develop our full potential in our many-sided lives. Yoga has become a way of life for many people all over the world in recent years. As a result, a scientific analysis of yoga postures is required. Recognizing posture is a difficult task due to the scarcity of datasets and the need to detect posture from a huge dataset. To address this issue, a large dataset containing five different types of yoga pose images was collected from the Kaggle online dataset. In the proposed work, the yoga pose is taken as input, the features were detected using, pre-trained models MobileNetV2 and DenseNet201. The Classifiers namely Support Vector Machine (SVM) and Random Forest (RF) are used for classification. By comparing the experimental results the performance of MobileNetvV2 with Random Forest yields better results when compared to other models.

Keywords - Support Vector Machine, Random Forest, MobileNetV2, DenseNet201.

1. INTRODUCTION

Yoga has become a well-known practise for maintaining people in good physical and mental health all around the world. [1] Asanas are just one part of an eight-part process that leads to enlightenment. They help the body to prepare for meditation [3]. A new Yoga Posture Classification (YPC) model using deep learning algorithms is proposed to classify the Yoga images using two different processes namely feature extraction and classification. In the proposed work, two Pre-Trained models MobileNetV2 and DenseNet201 are used as feature extractors to extract features from the input images. Then, the features are given to Support Vector Machine and Random Forest for classification of Yoga Posture. There are various yoga types some of them are used in this work; they are Downdog, Goddess, Tree, Plank and Warrior2.

1.1 YOGA POSTURE

Yoga may be intended as a way to calm the mind, but it also a great way to get in shape and drop some weight. Here are a few postures that can help lower your anxiety and the number on the scale. Hold each posture as long as you can, that may be 15-20 seconds at first, but each time you practice, hold the posture for a few seconds longer and make your way up to a minute if you can. Where appropriate do one side and repeat on the opposite side. There are various types of yoga posture. This work Downdog (Adho Mukha Svanasana), Goddess (Utkata Konasana), Plank (Phalakasana), Tree (Vrksasana) and Warrior2 (Virabhadrasana) are classified. Fig 1 shows the sample images of Yoga Posture.

DOWNDOG

Downward Facing Dog (Adho Mukha Svanasana) is the poster pose for yoga. The reason it has become the best-known asana is that it's so important in contemporary practice. It may be the first pose you learn as you begin a yoga practice. It is done many times during most yoga classes, particularly in Vinyasa yoga. It acts as a transitional pose and can be a resting position. Downward Dog is one of the poses in the Sun Salu Goddess Pose is considered a base pose as goddess pose variations can be derived from this pose. Goddess Pose helps boost energy in the body and hence can be included in flow yoga sequences.

GODDESS

Utkata Konasana (Goddess Pose), also referred to as the Fierce Angle Pose, is an empowering, intermediate standing level pose. The Goddess Pose (Utkata Konasana) can also be part of yoga sequences intended for hips, the chest as well as the groin. The practice of Goddess Pose, when part of a prenatal yoga sequence, helps to widen the uterus and prepare for delivery during pregnancy.

Goddess Pose is considered a base pose as goddess pose variations can be derived from this pose. Goddess Pose helps boost energy in the body and hence can be included in flow yoga sequences.

PLANK

Plank (*Phalakasana*) is one of the most foundational poses in yoga for lots of good reasons. Each time you visit this grounding pose you'll build abdominal strength and power in your arms and wrists. It also strengthen the muscles surrounding the spine, which improves posture. The increased strength and stamina you build in Plank will help you more challenging postures like Boat Pose, Dolphin Pose, and Side Plank Pose. The benefits of this posture aren't just physical. Plank Pose also builds mental endurance and deep focus. Holding the pose when your arms start to shake is a powerful reminder that you can tackle challenges both on and off the mat.

TREE

This posture replicates the graceful, steady stance of a tree. Unlike most yoga poses, the Tree Pose (Vrikshasana or Vrksasana) requires keeping our eyes open in order to maintain body balance. Avoid doing this posture if you are suffering from migraine, insomnia, low or high blood pressure (those with high blood pressure may do this pose but without raising their hands overhead, as this may further raise their blood pressure).

WARRIOR2

Warrior 2 or Virabhadrasana 2, is a grounding and lunging standing yoga pose. Derived from the root words Vira (warrior), Bhadra (friend) and Asana (pose), it increases mind and body awareness as it builds stamina, power and strength. Warrior2 pose is sometimes seen as anatomically difficult because you have to maintain balance and work on breathing as your legs go in different directions. Hence, a great deal of concentration is required to focus on your front leg being bent while the back leg is straight. You also have to direct your pelvis to ensure its pointing forward and level.

Downdog **Goddess** **plank**

Tree **Warrior2**

Fig. 1 Sample images of Yoga Posture

In the proposed work, the pre-trained models namely, MobileNetV2 and DenseNet201 are used for feature extraction and the classifiers namely Support Vector Machine (SVM) and Random Forest (RF) are used to classify the features into the respective classes. The block diagram of the proposed work is shown in Fig. 2.

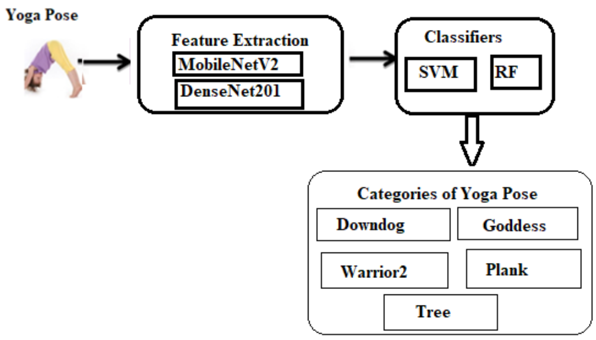

Fig. 2 Block Diagram of Proposed Work

2. LITERATURE REVIEW

The yoga participants conducted the research [1]. The data was gathered from yoga studios. In this study, 100 observations were collected from yoga participants in Ammapet, Salem. Any serious yoga practise necessitates the use of mindfulness. Yoga, when done correctly, clears the mind of all distracting ideas from the outer world, bringing one to a state of inner serenity [2]. 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 Support Vector Machine (SVM) approach seeks to find the optimal separating hyper plane between classes by focusing on the training cases that are placed at the edge of the class descriptors. These training cases are called support vectors. Training cases other than support vectors are discarded. In this way not only is an optimal hyperplane fitted but also less training samples are effectively used thus high classification accuracy is achieved with small training sets [3,4].

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].

[14] did research on Deep Learning (DL), Reinforcement Learning (RL), and their combination (Deep RL) promise to revolutionize Artificial Intelligence (AI). The growth in computational power accompanied by faster and increased data storage and declining computing costs have already allowed scientists in various fields to apply these techniques on datasets that were previously intractable for their size and complexity. This review article provides a comprehensive survey on the application of DL, RL, and Deep RL techniques in mining Biological data. In addition, they compare performances of DL techniques when applied to different datasets across various application domains. Finally, we outline open issues in this challenging research area and discuss future development perspectives.

[15] developed a Yoga self-training system that uses a Kinect depth camera to aid in correcting postures while practising Yoga for 12 different asanas. It, on the other hand, employs manual feature extraction and creates unique models for each asana. Delegate features, like as a human skeleton, must be extracted in order to describe human postures. Various skeletonization strategies, such as thinning and distance transformation, have been documented in the literature.

Using deep learning techniques such as convolutional neural networks and transfer learning, the system detects the yoga posture from both images and videos. The model was trained and the prediction accuracy was evaluated using ten different yoga asana. One of the pre-trained models, VGG16, was utilised as a feature extractor, and a deep neural network classifier was employed to identify the output classes. In this study, a prediction accuracy of 85 percent yields [9-13].

3. FEATURE EXTRACTION

3.1 MobileNetV2

It is originally designed for mobile device by google. MobileNetv2 is a fine pre-trained model that delivers the high output accuracy. MobileNetv2 is built upon the idea of MobileNetv1. It uses the depth wise separable convolution as efficient building blocks [6]. MobileNetv2 has two important features they are linear bottleneck between the layers and shortcut connections (residual block in the network) between the bottleneck. There are two types of blocks in MobileNetv2. They are residual block with stride of 1 and other block is stride of 2 with downsizing. Fig. 3 shows the mobileNetv2 convolutional blocks. The Table 1 shows the bottleneck of mobilenetv2.

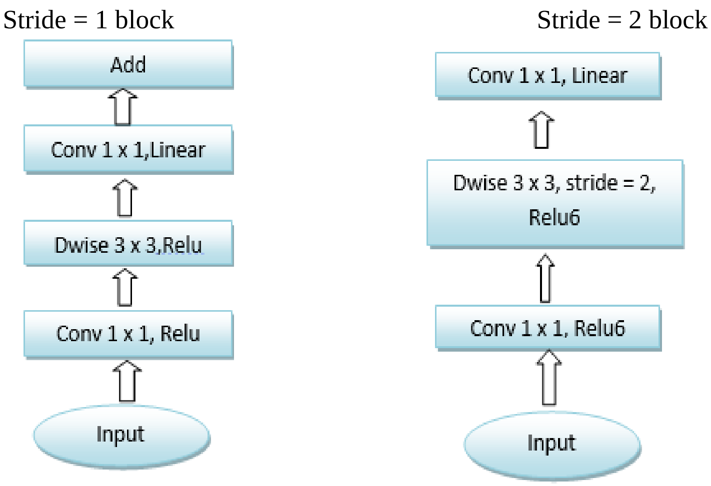

Fig. 3 MobileNetv2 Convolutional Block

Both type of the block has three layers.

- First layer has 1 x 1 convolution with Relu6 as the activation function.
- Second layer has the depth wise convolution.
- The third layer has another 1 x 1 convolution, but without any non-linearity.

Table 1 Bottleneck of MobileNetv2

Input	Operator	Output
h x w x k	1 x 1 Conv2d, ReLU6	*h x w x (tk)*
h x w x tk	3 x 3 dwise s=s ReLU6	$\frac{h}{s} \times \frac{w}{s} \times (tk)$
$\frac{h}{s} \times \frac{w}{s} \times tk$	Linear 1 x 1 Conv2d	$\frac{h}{s} \times \frac{w}{s} \quad k^1$

Where t is expansion factor, c is the number of output channels, n is the repeating number, s is the stride, for spatial convolution 3 x 3 kernels are used.

3.2 DenseNet201

In a DenseNet201 architecture, each layer is connected to every other layer, hence the name Densely Connected Convolutional Network. For L layers, there are L(L+1)/2 direct connections. For each layer, the feature maps of all the preceding layers are used as inputs, and its own feature maps are used as input for each subsequent layers. DenseNets essentially connect every layer to every other layer. This is the main idea that is extremely powerful. The input of a layer inside DenseNet is the concatenation of feature maps from previous layers. Fig. 4 shows the DenseNet201 architecture

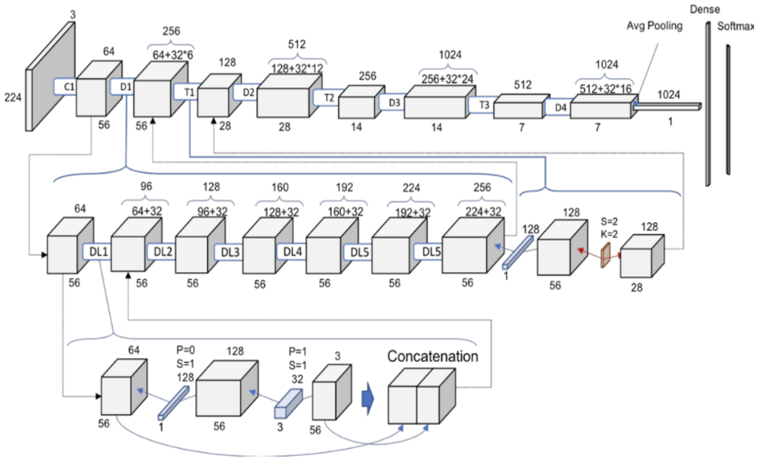

Fig. 4 DenseNet201 Architecture

4. CLASSIFIERS

In this proposed work SVM and RF are the machine learning techniques used to classify the features.

4.1 Support Vector Machine

SVM is a statistic machine learning technique that has been successfully applied in the pattern recognition area and, is based on the principle of structural risk minimization. SVM builds a straight model to assess the choice capacity utilizing non-direct class limits in light of help vectors. [12] SVM learns an optimal separating hyper plane from a given set of positive and negative.

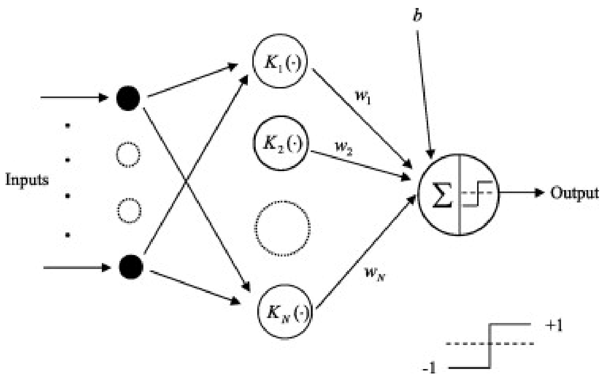

Fig. 5 SVM architecture

Fig. 5 shows the architecture of the SVM. It maps the input patterns into a higher dimensional feature space through some nonlinear mapping chosen a priori. A linear decision surface is then constructed in this high dimensional feature space. Thus, SVM is a linear classifier in the parameter space, but it becomes a non-linear classifier as a result of the nonlinear mapping of the space of the input patterns into the high dimensional feature space.

4.2 Random Forest

Random forest (RF) consists of a large number of individual decision trees that operate as an ensemble. Each distinct tree in the random forest spits out a class prediction and the class with the most votes becomes the model's prediction. The fundamental concept behind random forest is a simple but powerful one, the wisdom of crowds. A large number of relatively uncorrelated trees that operating as a committee will outperform any of the individual constituent models.

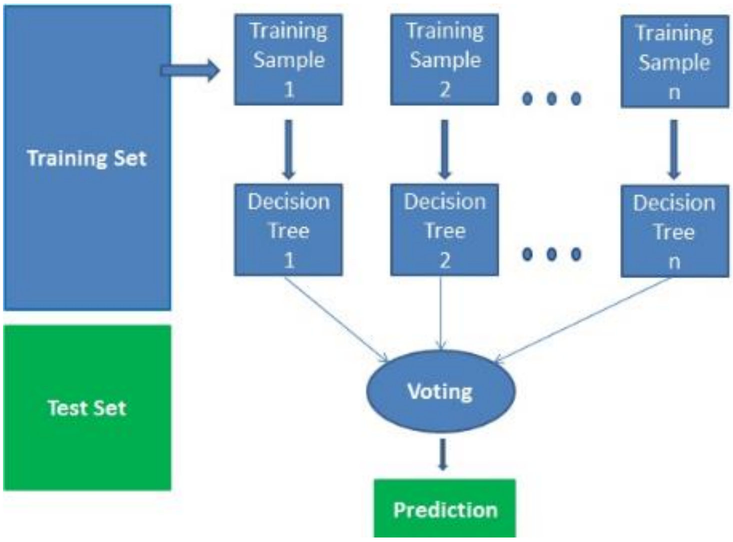

Fig. 6 Classification Process of Random Forest

5 PROPOSED WORK

5.1 Experimental Results

By analyzing the effectiveness of the network and by utilizing the Pre-Trained model namely MobileNetV2 and DenseNet201 as a feature extractor, the weights are transferred to the proposed system. The optimal setup values vary while fine tuning the deep networks. In the Pre-Trained models the FC layer followed by an output layer was selected and was replaced by the SVM, RF and GNB classifier.

5.2 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.

5.3 *Pre-Trained MobileNetV2 as a Feature Extractor*

The input layer adopts an image in the size of 224 x 224 x 3 and the output layer is the softmax prediction on 1000 classes. From the input layer to the end is the max pooling layer by 7x7x 1280 which is regarded as the feature extraction of the model, while the rest of the network is presented as the classification of the model. In this work, MobileNetV2 is applied to extract the features and we conclude at 1280 features for a single image.

5.4 *Pre-Trained DenseNet201 as a Feature Extractor*

DenseNet201 takes the input image of size 112 x 112 and the output layer is the SoftMax prediction for 1000 classes. From the input layer to the last is the Avg pooling layer by 7 x 7 dimensions which is considered as the feature extraction part of the model, while the rest of the network is considered as the classification part of the model. In this work, DenseNet is used to extract the features and we arrived at 1024 features for a single image.

6. PERFORMANCE MEASURES

The overall performance of F-Score, precision and recall of Yoga Posture classification is shown in Fig. 7 and 8. The overall accuracy of SVM is shown in Table 2 and the overall accuracy of RF is shown in Table 3. The Yoga Posture classification 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

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

The performance of MobileNetV2 with RF gives better performance when compared to other models.

6.1 Classification using SVM

SVM is trained to identify the features. For training 1080 images were trained with the feature vectors, each of 1280 dimension for MobileNetV2 and 1024 dimension for DenseNet, are extracted from the images. The training process analyses the YPC training data to find an optimal way to classify Yoga images into its respective classes namely, Downdog (category 0), Godess (category 1), Plank (category 2), Tree (category 3) and Warrior2 (category 4). The derived support vector is used to categorize the Yoga input images. For testing 470 feature vectors are given as input to the SVM model and the distance between each of the feature vector and the SVM hyperplane is calculated. The average distance is calculated for each model. The sssstages of YPC are decided based on the maximum distance. The classification results are analyzed based on four measures namely Precision, Recall, F-score and Accuracy. Table 2 shows the accuracy of SVM with MobileNetV2 and DenseNet201.

Table 2 Accuracy of SVM

Class	Accuracy	
	MobileNetV2	DenseNet201
Downdog	58.54	58.54
Goddess	84.38	78.12
Plank	58.70	78.26
Tree	68.97	62.07
Warrior 2	89.36	72.34

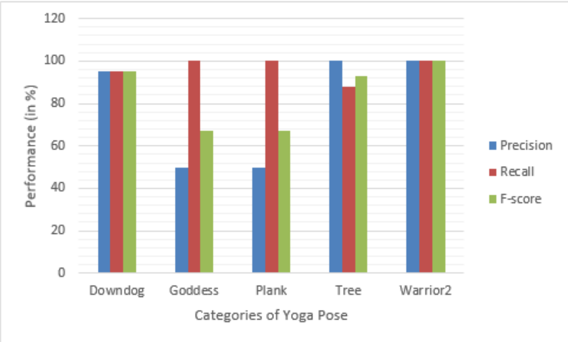

Fig. 7 Performance of MobileNetV2 with SVM

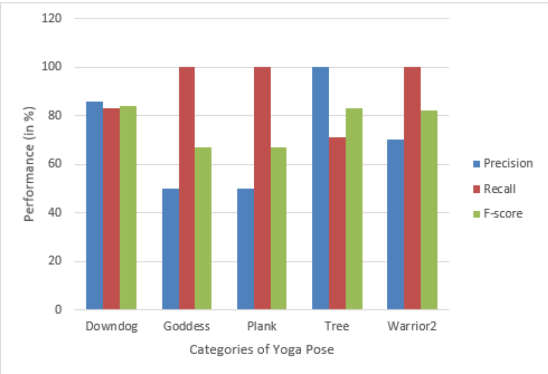

Fig. 8 Performance of DenseNet201 with SVM

6.2 Classification using Random Forest

The training process analyses the training data to find the decision tree to classify the images into five categories. Random forest is structured for a complete learning procedure for categorizing with a set of decision trees that grow randomly by selecting the sample data. A nonlinear decision tree is applied to discriminate the various stages. Random Forest is trained to ascertain image features. The bootstrap procedure is followed for each training set; the samples are selected randomly. For training 1080 images were trained, 1280 dimension for MobileNetV2 1024 dimension for DenseNet201 are extracted from the images. At each point, a new subset is generated; current tree is drawn by replacement of vectors. This is called as out-of-bag. For testing 470 feature vectors are given as input to the Random Forest model. While testing, predictions are arrived by finding the average of the study of each decision tree. Table 3 shows the accuracy of RF with MobileNetV2 and DenseNet201.

Table 3 Accuracy of RF

Class	Accuracy	
	MobileNetV2	DenseNet201
Downdog	87.80	87.80
Goddess	78.12	78.12
Plank	84.78	86.96
Tree	89.66	79.31
Warrior 2	91.49	85.11

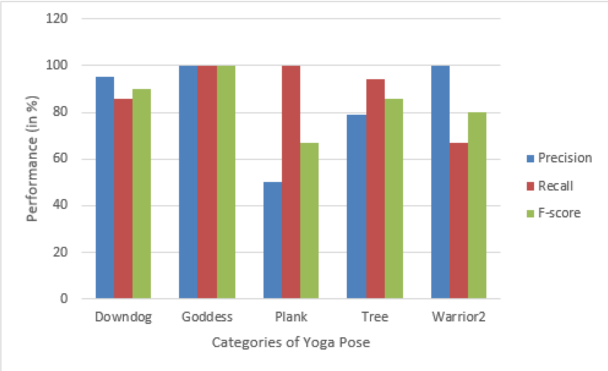

Fig. 9 Performance of MobileNetV2 with RF

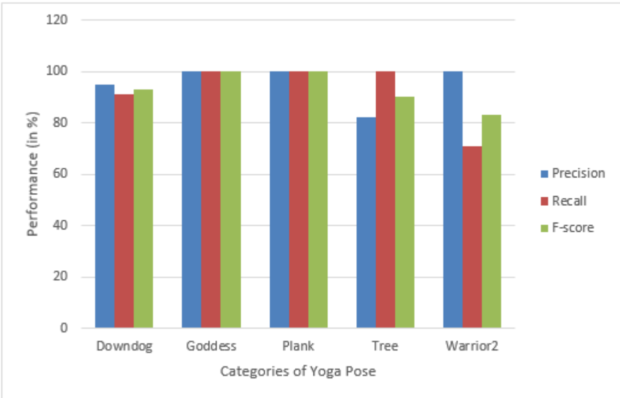

Fig. 10 Performance of DenseNet201 with RF

7. CONCLUSION

In the proposed work, yoga postures are classified using SVM and RF model using the MobileNetV2 and DenseNet201 features. In this paper, machine learning techniques such as Support Vector Machine (SVM) and Random Forest (RF) are proposed. The performance were extracted from the Yoga Posture images using the two feature extraction techniques namely MobileNetV2 and DenseNet201 models which is the pre-trained convolutional neural networks. The result for the proposed work gives satisfactory results, that MobileNetV2 with RF model gives good results when compared to the other model.

REFERENCES

1. 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.
2. 2,100 Asanas: The Complete Yoga Poses, (2015), By Daniel Lacerda
3. Mercier, G., and Lennon, M., (2003), Support vector machines for hyperspectral image classification with spectral-based kernels, Proceedings in IEEE International Geoscience and Remote Sensing Symposium, Vol. 1, pp. 288-290.
4. Yoga for Your Type: An Ayurvedic Approach to Your Asana Practice, By David Frawley, Sandra Summerfield Kozak, 2017
5. Mahmud, M., Kaiser, M. S., Hussain, A., and Vassanelli, S., (2018), Applications of Deep Learning and Reinforcement Learning to Biological Data, IEEE Transaction on Neural Networks and Learning System, Vol. 29(6), pp. 2063-2079.
6. Asanas: 608 Yoga Poses, By Dharma Mittra, 2003.
7. 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.
8. 7. Yadav, S.K., Singh, A., Gupta, A. et al. Real-time Yoga recognition using deep learning. Neural Comput&Applic 31, 9349– 9361 (2019)
9. 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: D

10. ExNET: Deep Neural Network for Exercise Pose Detection
11. 10. Lowe, D. G., (2004), Distinctive image features from scale invariant key points, International Journal of Computer Vision, Vol. 60(2), pp. 91-110.
12. 11. Josvin Jose, Shailesh S., (2021), Yoga Asana Identification: A Deep Learning Approach, IOP Conf. Series: Materials Science and Engineering 1110 (2021) 012002.
13. 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 microsoft kinect. Pp. 668–673.
14. Chen HT, He YZ, Hsu CC, (2018) Computer-assisted yoga training system. Multimed Tools Appl 77:23969–23991. https://doi.org/10.1007/s11042-018-5721-2
15. Mala, K., (2020), Importance of Yoga and Spirituality in Modern Lifestyle, International Journal of Engineering Research and General Science Volume 8, Issue 1, ISSN 2091-2730, pp. 13-39