e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:12/December-2021 Impact Factor- 6.752 www.irjmets.com

YOGA POSE DETECTION AND CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

Utkarsh Bahukhandi*1, Dr. Shikha Gupta*2

*1Student, Department Of Information Technology, Maharaja Agrasen Institute Of Technology, Rohini, Delhi, India.

*2Assistant Professor, Department Of Information Technology, Maharaja Agrasen Institute Of Technology, Rohini, Delhi, India.

ABSTRACT

Yoga is an ancient art with a long history associated with India. It helps in making a person physically fit and provides mental peace at the same time. With the introduction of Covid-19, it is difficult to perform yoga in classes and if performed without guidance it may cause some serious injuries. Here we develop a system that identifies different yoga poses performed by users. The system uses open-source data containing 6 different yoga poses videos performed by 15 different volunteers. The system has two phases first to extract the data points data from the video dataset using the media pipe pose estimation library and the second phase is preprocessing the obtained data, training, and testing the data using classification-based machine learning algorithms. The machine learning algorithm used is logistic regression, support vector machine classifier, random forest classifier k nearest neighbors classifier, and naïve Bayes classifier. The system achieves an accuracy score of 94%. The system is developed to work on images, static videos, and live videos with a threshold value so that below a certain score it does not accept the solution.

Keywords: Pose Estimation, Computer Vision, Machine Learning, Classification, Yoga, Health, Classifiers, Logistic Regression, Support Vector Machine, KNN, Random Forest, Naïve Bayes.

I. INTRODUCTION

Yoga is an art that originated in India a long time back. It helps boost the physical health of a person and cleanses the body, mind, and soul of a person. Yoga can cure many diseases without any need for medicines With the introduction of Covid-19 humans understood one thing that it is health which is more important than anything else in this world and the situation is very difficult for everyone as there is bad news from everywhere in the world which makes disturbance to mental peace of everyone, hence yoga acts as a perfect recipe for this situation. For detection of yoga poses in any system human pose estimation is required. Human pose estimation is the computer vision problem in which the human pose of any person is detected when the camera receives an image of a human in front of it. The detection of human poses is done using key points detection. These key points are the major points of the human body which include the nose, eyes, mouth, etc. There are two different ways to detect these key points

- **Top-down approach:** The major work behind the top-down approach is that it first finds bounding boxes that include every person in the frame. Next for every bounding box, it finds out the joint position of the person in the box. Hence every bounding box has its joints associated with it.
- **Bottom-up approach:** This approach is the opposite of the top-down approach. In this first, all the joints that are available in the image are found, and then joints corresponding to every bounding box are separated to classify them according to each person.

Some major work is done in the yoga poses detection using the human pose estimation field. Sruthi Kothari worked on a method that uses deep learning mainly convolutional neural networks for classifying yoga postures in images, the dataset employed consists of 1000 images distributed for 6 different yoga poses and obtained 85% accuracy for this work. Muhammad Usama Islam proposed a different method by finding the joint points of the human body and using then calculating the joint angles of the human body to estimate asanas or poses and accuracy is found out using Microsoft kinetic. So if angles fall below 97% then the pose is shown as no pose detection.

Mediapipe pose estimation: This is a pose estimation method developed by researchers of google and operates on the blaze fast model for the pose detection method. It is a fast model and performs at a 24FPS rate

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:12/December-2021 Impact Factor- 6.752 www.irjmets.com

and hence is perfect for live video pose estimation. Blazepose model returns 33 key points or landmarks from the given image in which a human is detected. These points are major joint points of a human body and the points returned are 3-D coordinates with a visibility value. For a nonvisible joint, it predicts the coordinates of the joint using the concept of Leonardo's Vitruvian man and hence the midpoint of person's hip, the radius of a circle consisting of human and inclined line angle connecting shoulder and hip's midpoint is predicted.

Classification in machine learning is the supervised machine learning methodology that works on obtaining certain values as output for several features. These output values are known as target variables and their number is always greater than 1. The algorithms are mathematical functions that work on cost and error obtained for every data we provide and train in such a way as to minimize the error using a specific learning rate and gradient descent. Examples of some classification-based machine learning algorithms are logistic regression, support vector machine classifier, random forest classifier, k nearest neighbor classifier, naïve Bayes classifier, etc.

II. METHODOLOGY

The system comprises of two phases first is the training phase in which training of machine learning models happens and next is the testing phase in which testing of trained models and evaluation of their performances happens.

Training Phase

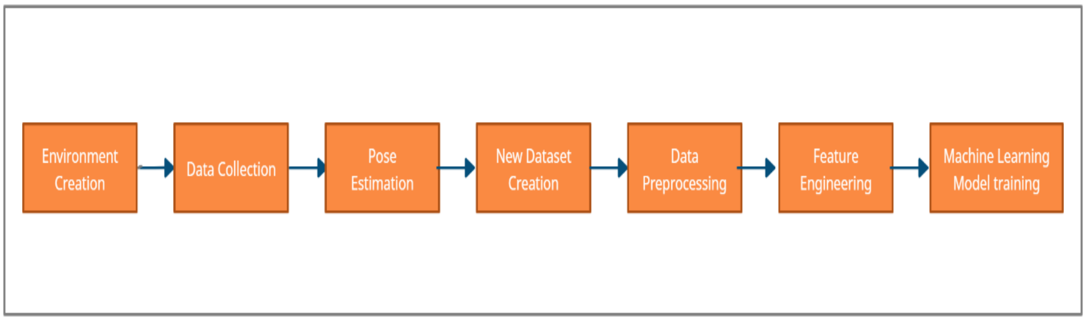

Figure 1: Workflow of the training phase

Firstly the environment for the system is created and all the necessary libraries are installed. The libraries that are necessary for the system are- NumPy for numerical calculation, pandas for the dataset creation and data operations, seaborn for data visualization, sklearn for machine learning algorithms instances, OpenCV for computer vision tasks, media pipe for pose estimation.

After environment creation data collection is done, hence open-source data is used from Kaggle where 6 yoga poses were performed by 15 different volunteers both male and female. The data is in video format hence every frame of video is taken and fed to the next part of the system.

Then the images are fed to the media pipe for pose estimation it then detects the 33 key points on the given frame and provides the value of 3-D coordinates of these key points with the visibility value. These 33 key points then form the basis of the new dataset and every key point coordinate and its visibility becomes the feature of a new dataset.

Then data preprocessing methods are applied to clean the data and make everything suitable for the machine learning models. For preprocessing of data normalization is applied to the dataset. Normalization is the process in which we make every value present in the dataset lie between zero and one. The primary reason for doing so is that certain machine learning models require normalized data to perform efficiently and provide good results. Then feature engineering is employed to obtain new features from the existing features in the dataset. To do so conversion of key points to vectors is done. These vectors are 3-dimensional body parts and using these vectors joint angles are calculated.

Finally, all the processed data is then fed to the classification-based machine learning algorithms to train them.

Testing Phase

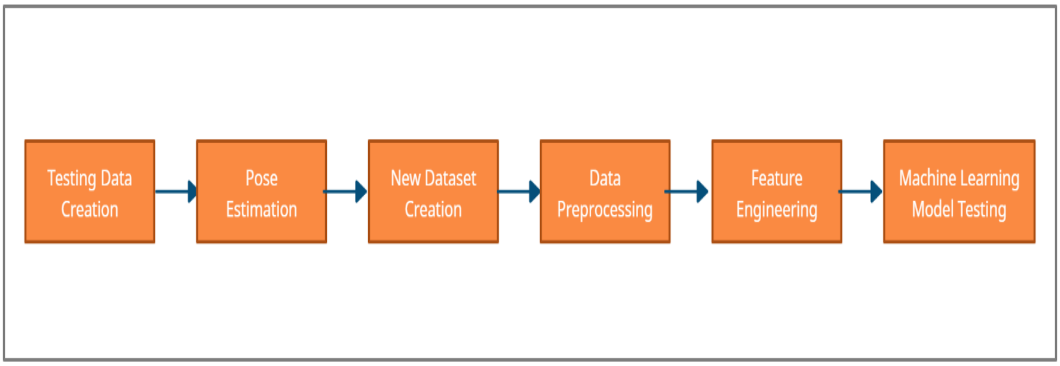

Figure 2: Workflow of the testing phase

Firstly testing data is to be obtained/created, this data must be new and should not consist of any redundant data from the training phase. To obtain testing data yoga poses video is created by a different volunteer who is present at a different location. These videos are taken as test data and the equipment used for this is an HD 720P Web Camera. The frames of each video are then sent to the media pipe pose estimation model which finds the 33 key points onto each frame.

Then similar preprocessing and feature engineering is done on the testing data which was done on training data to make testing data similar to training data. The reason for this is that machine learning models understand the training data.

Finally, the output is obtained for the test data and then fed to the trained model to evaluate its performance based on its predictions.

III. MODELING AND ANALYSIS

Reading and Understanding the data: Video frames are fed to the pose estimation model and key points dataset is obtained.

	NOSE_x	NOSE_y	NOSE_z	NOSE_visiblity	LEFT_EYE_INNER_x	LEFT_EYE_INNER_y	LEFT_EYE_INNER_z	LEFT_EYE_INNER_visiblity	LEFT_EYE_x	LEFT_
0	0.432962	0.586089	-0.162905	0.999993	0.437066	0.577206	-0.153582	0.999997	0.439082	0.
1	0.492706	0.326348	-0.332448	0.999912	0.495855	0.314325	-0.318994	0.999875	0.498365	0.
2	0.537991	0.324800	-0.189326	0.999641	0.542028	0.311678	-0.172213	0.999432	0.544572	0.
3	0.524368	0.595201	0.014085	0.999924	0.519550	0.590234	0.027053	0.999872	0.518649	0.
4	0.556475	0.590921	-0.216842	0.999965	0.559876	0.580551	-0.207372	0.999977	0.561862	0.
5	0.501702	0.324782	-0.329311	0.999852	0.504862	0.311949	-0.316156	0.999785	0.506924	0.
6	0.566081	0.584218	-0.072731	0.999886	0.562522	0.574308	-0.058597	0.999791	0.561964	0.
7	0.447344	0.345254	-0.116424	0.999744	0.451519	0.335693	-0.103173	0.999508	0.453775	0.
8	0.604666	0.460148	-0.469899	0.999992	0.612649	0.465145	-0.455835	0.999973	0.612634	0.
9	0.280518	0.732486	-0.093930	0.998035	0.274698	0.740138	-0.084654	0.995515	0.274604	0.

10 rows × 133 columns

Figure 3: Understanding the data

Figure 3 above shows the overview of our new dataset that we have just obtained from the videos dataset. The figure exactly describes how our dataset looks like and what are the features and attributes we are working with. The dataset has 133 columns.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:12/December-2021 Impact Factor- 6.752 www.irjmets.com

Data Preprocessing:

```
df.target.value_counts()

Vriksh        16910
Bhuj          13830
Padam         12693
Shav          12289
Trik          12147
vriksh         5311
Tad            4978
Tadasan        4366
Tada           4084
padam          2667
Trikon         2644
tadasan        2513
padmasan       2511
```

Figure 4: Target unique value counts

Figure 4 shows that though we had only 6 target values our dataset still has more than 6 target variables hence it needs to be processed. So all the target variables are taken and based on their initial three letters the target variables are given names.

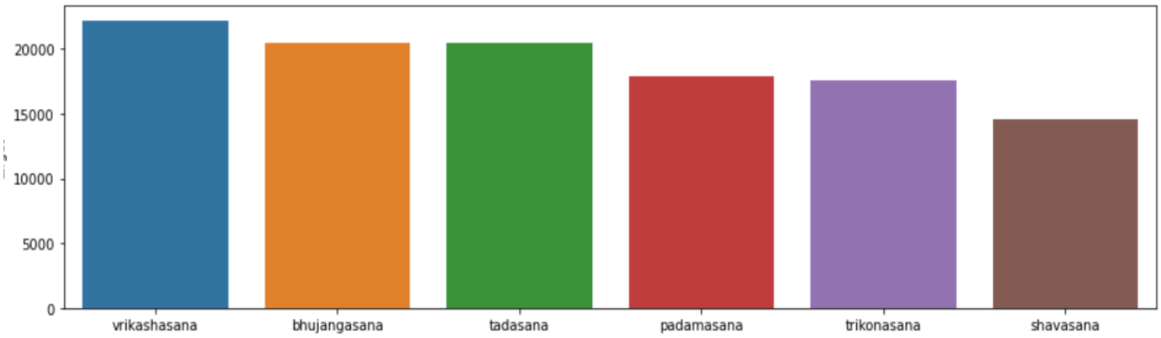

Figure 5: Value count of every target variable

Figure 5 clearly shows that after preprocessing there are now only 6 target variables. Also, it shows that the values are comparable to each other, hence we do not require sampling of data because it will not have any skewness. After this, we normalize the data to scale it between 0 and 1.

Feature Engineering: As only data points will not be sufficient for pose classification as it may be possible that the position of body parts is the same but the pose is different so there is a need for the introduction of features that eliminates complete dependency on the key points. For this joint angle to be calculated, we first make a vector of every body part using 2 key points and then calculate the angle of two vectors using the formula shown in figure 6.

$$\theta = \cos^{-1}\left(\frac{\boldsymbol{A}.\boldsymbol{B}}{|\boldsymbol{A}||\boldsymbol{B}|}\right)$$

Figure 6: Angle between two vectors

Model training: As the analysis of data is complete, modeling of data must be done using a classification-based machine learning algorithm. So the system uses 5 different machine learning algorithms namely- logistic regression, support vector machine classifier, random forest classifier, k nearest neighbors classifier, and naïve Bayes classifier. For the evaluation of machine learning models, accuracy is used as an evaluation criterion.

IV. RESULTS AND DISCUSSION

Accuracy best predicts the successful classification out of a total number of samples and determines the model's ability to correctly predict the target value. The logistic regression classifier is having the best accuracy score among all the classifiers employed with an accuracy score of 94%. Random forest classifier performs worst among all with an accuracy score of 89%. For an image/static video/live video after pose prediction, a threshold value is set up below which the system generates the output as no pose detected. The confidence level required for pose detection is 97%.

Table 1. Comparison of the score of machine learning models

SN.	Machine Learning Model	Accuracy	Precision	Recall	F-1 Score
1	Logistic regression	0.94	0.94	0.94	0.94
2	SVM classifier	0.93	0.94	0.93	0.93
3	Random Forest classifier	0.89	0.90	0.89	0.89
4	Knn classifier	0.93	0.94	0.93	0.93
5	Naïve Bayes classifier	0.91	0.93	0.89	0.91

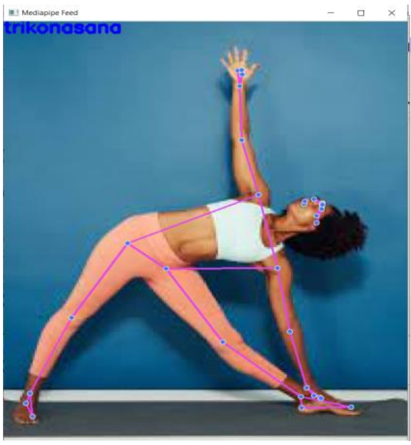

Figure 7: Pose detection Successful

Figure 8: Pose detection unsuccessful

V. CONCLUSION

In this study, a yoga pose classifier was successfully developed which works perfectly on images, static video, and live video of any user. The study starts from environment creation and proceeds with data collections from open data sources. Mediapipe pose estimation library is used for human pose estimation which returns body key points, these data points form the basis of a new dataset. Then data preprocessing takes place in which target variables are changed. After this normalization of data occurs for better performance of machine learning algorithms and finally feature engineering of features starts where various joint angles of the body are calculated using the formula shown in figure 6. As the data is completely preprocessed data is finally fed to machine learning models. Evaluation of these models is done on test data and is compared based on accuracy score. Logistic regression classifier achieves a maximum score of 94% among all classifiers. For classification a threshold value is used which is set at 97% below which no pose detected is given as output to the user.

VI. REFERENCES

[1] Guddeti RR, Dang G, Williams MA, Alla VM (2018) Role of Yoga in cardiac disease and rehabilitation. J Cardiopulm Rehabil Prev. https://doi.org/10.1097/hcr.0000000000000372

[2] Dr. Gaba, Importance of Yoga in the Pandemic, https://sarvodayahospital.com/blog/importance-of-yoga-during-this-pandemic

[3] Classification of yoga pose using machine learning techniques Author: J. Palanimeera, K. Ponmozhi https://doi.org/10.1016/j.matpr.2020.08.700

[4] Ajeet Ram Pathak, Manjusha Pandey, Siddharth Rautaray, Application of Deep Learning for Object Detection, Procedia Computer Science, Volume 132, 2018, Pages 1706-1717, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.05.144. (https://www.sciencedirect.com/science/article/pii/S1877050918308767)

[5] Shruti Kothari 2020 Yoga Pose Classification Using Deep Learning Ph.D. thesis SAN JOSE STATE UNIVERSITY

[6] Y. Agrawal, Y. Shah, and A. Sharma, "Implementation of Machine Learning Technique for Identification of Yoga Pose," 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), 2020, pp. 40-43, DOI: 10.1109/CSNT48778.2020.9115758.

[7] Narayanan, S. Sankara, and Misra, Devendra Kumar and Arora, Kartik and Rai, Harsh, Yoga Pose Detection Using Deep Learning Techniques (May 10, 2021). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021, Available at SSRN: https://ssrn.com/abstract=3842656 or http://dx.doi.org/10.2139/ssrn.3842656

[8] Kumar, Deepak & Sinha, Anurag. (2020). Yoga Pose Detection and Classification Using Deep Learning. International Journal of Scientific Research in Computer Science Engineering and Information Technology. 10.32628/CSEIT206623.

[9] Nandakishore Joshi August 17, 2020, REALTIME 2D YOGA POSE ESTIMATION WITH CODE WALKTHROUGH, Medium, https://nandakishorej8.medium.com/realtime-2d-yoga-pose-estimation-with-code-walk-through-cfd69262d356

[10] Syed Abdul Gaffar Shakhadri — May 18, 2021, Pose Estimation using OpenCV, Analytics Vidhya, https://www.analyticsvidhya.com/blog/2021/05/pose-estimation-using-opencv/

[11] Cramer, Jan Salomon. "The origins of logistic regression." (2002): 4

[12] Bazarevsky, Valentin, et al. "BlazePose: On-device Real-time Body Pose tracking." arXiv preprint arXiv:2006.10204 (2020).

[13] Pauzi A.S.B. et al. (2021) Movement Estimation Using Mediapipe BlazePose. In: Badioze Zaman H. et al. (eds) Advances in Visual Informatics. IVIC 2021. Lecture Notes in Computer Science, vol 13051. Springer, Cham. https://doi.org/10.1007/978-3-030-90235-3_49