YOGA TRAINER WITH AI

Shivani Balakrishnan[1], Shriya S[2], Srajana[3], Vaishnavi N[4] , Dr. Kavitha C[5]
[1234] *UG Students, Computer Science, and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India*
[5] *Professor & Head, Computer Science, and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India*

ABSTRACT

Yoga is beneficial for everyone of any age's physical and mental well-being. To avoid damaging the bones, muscles, and ligaments, it is extremely important to execute yoga poses correctly when practicing alone. As a result, providing feedback to the performance without a live instructor is possible by using artificial intelligence and machine learning in conjunction with picture processing. In addition to correcting users who perform the yoga posture improperly, the suggested system is designed to advise users on how to do it correctly on real-time and there is also a pose classification feature that can classify the yoga poses based on the images of the final yoga pose which is provided by the user. The feedback, which comes in text and audio formats, can assist the practitioner avoid injuries while also maximizing the advantages of doing the yoga posture. Various images from the internet were used to create the data set of various yoga poses. The data points are extracted from each webcam image with the aid of OpenCV and media pipelines. This is now fed into a deep learning model that employs convolutional neural networks (CNN), which finds pose errors and calculates the error percentage before providing the user with the necessary feedback in the form of text or audio desired output with a classification accuracy of roughly 95%.

Keywords: - *Human pose estimation, Media pipes, Activity recognition, real-time pose detection, Yoga.*

1. INTRODUCTION

Yoga, which has its roots in India, is a type of physical, mental, and spiritual exercise that promotes balance between the body and the mind. It is one of the oldest sciences in existence and has the potential to enhance both physical and psychological health. Improved flexibility, energy, sleep, posture, muscle strength, circulatory and cardiac health, as well as relief from chronic symptoms are all advantages of regular yoga practice. However, it has been demonstrated that poor yoga postures can result in fractures, strains, sprains, joint dislocations, nerve damage, and stroke.

Artificial intelligence has benefited people in many areas of life. Whether it be in the form of a medical application, mobility, healthcare, or retail, technology has impacted practically every business. AI is now being used in the fitness sector. In general, when it comes to fitness, many individuals consider food, cardio, the gym, and yoga, which is by far the most popular alternative. Since many people lack the means to hire a yoga instructor or guide, artificial intelligence can step in and provide guidance on the precision and timing of poses. This project indicates whether you are performing the stance correctly, how long you can hold it for, and your best performance in the pose.

Artificial intelligence (AI) refers to a machine's capacity to mimic or improve human intelligence, such as deductive reasoning and experience-based learning. Although it has long been used in computer programs, artificial intelligence is now used in a wide range of other goods and services. For instance, some using artificial intelligence software, digital cameras can identify the objects in an image. Smart electric grids are one of the numerous futuristic applications of artificial intelligence that scientists think will emerge. To answer real-world issues, AI employs methods from probability theory, economics, and algorithm design. Additionally, the field of AI incorporates linguistics, psychology, mathematics, computer science, and more. Mathematics gives methods for modelling and resolving the ensuing optimization problems, whereas computer science offers tools for creating and constructing algorithms.

Machine learning (ML) allows computers to do particular tasks and make judgements more effectively with the aid of data and experience. Probability theory and statistics are used in this by ML. Machine learning uses algorithms to examine data, learn from it, and make inferences without the use of explicit programming.
Machine learning algorithms are usually separated into supervised and unsupervised groups. While supervised algorithms can apply what they have learned to new data sets, unsupervised algorithms can draw conclusions from datasets. Finding both linear and non-linear relationships in a given set of data is the aim of machine learning techniques. This is achieved via the statistical methods used to train the algorithm to classify or predict from a dataset.
Deep learning is a kind of machine learning that use multiple layers of artificial neural networks to achieve cutting-edge accuracy in language translation, object detection, and speech recognition. A key component of driverless automobiles is deep learning, which enables machine analysis of massive volumes of complicated data, such as identifying faces in pictures or videos.

2. PROBLEM DESCRIPTION

The yoga posture coaching system utilizing a web-based CNN machine learning technique is proposed in the following project. It intends to create a platform powered by artificial intelligence that assists users in practicing yoga in the right and ordered manner. There will be a variety of yoga poses and a requirement for internet connectivity.

3. RELATED WORKS

Various companies have developed technology products for the sports and exercise industry. For instance, [1] created NADI X-Smart Yoga Pants that work with a mobile app to guide exercise forms. Another popular product is [2], a movable gym mirror that allows users to view their exercises and reflection simultaneously.
[3]SmartMat designed an intelligent yoga mat with sensors that detect posture pressure and give real-time feedback via a mobile app. [4]YogaNotch created a wearable device that gives audio feedback on alignments during home yoga practices.
Yoga posture identification with a smartphone application, however, is challenging and requires close user monitoring in order to provide the right feedback [5]. The suggested yoga mat with a pressure sensor in [6] was insufficient for correcting posture throughout the body. For exercising while reflecting, the previous products produced in [1,3,4] are insufficient. Additionally, [2], the product that included a reflection monitor, still has problems because it only includes a camera without any body-posture detection or correction. The cost of using MIRROR, which includes product costs and monthly training class fees, may also be a deterrent for some.
The learning of 3D landmark points from a single image was applied by Nagalakshmi et al. [7] utilising skinned multi-person linear (SMPL) and an encoded architecture of classification models. kNN, SVM, and other deep learning models including AlexNet, VggNet, and ResNet were among these models. The proposed dataset, which was built on a collection of 13 different yoga postures, was used with each model. A half-camel pose, a standing half-forward bend, a bridge posture, a seated forward bend, a child posture, a corpse stance, a mountain posture, a tree posture, a triangle posture, and a twisted posture are all included in the dataset. The accuracy of classification was 83% when it was examined.
In numerous research that make use of diverse techniques, like the OpenPose algorithm[8] for posture estimation and machine learning and deep learning methods for posture recognition, yoga classification is crucial. [9] classified yoga poses for several people in real-time utilising posture estimation with a 3D pose from an RGB camera. They concentrated on 12 sun salutation asana poses from yoga in [9]. The poses were recorded by a webcam. With the help of machine learning models like support vector machine (SVM), k-nearest neighbors (kNN), nave Bayes, and logistic regression, the method involved building a skeleton, extracting features, and classifying the sun salutation yoga postures. It achieved a 96% accuracy rate.
A computer-assisted self-training system that detected the downward-facing dog, warrior 3 stance, and tree pose while also extracting the user's body contour was created with the use of [10]. They were only 82.84% accurate overall in the studies they completed. The previous research provided a self-training system that estimated the yoga posture by extracting the visual features from image data to analyse the 12 yoga postures, such as the tree posture, warrior 3 posture, warrior 2 posture, warrior 1 posture, downward-facing dog posture, extended hand-to-big-toe posture, chair posture, full boat posture, cobra posture, side plank posture, plank posture, and lord of the dance posture[11]. They recorded the user body map in order to extract the practitioner's body's feature points and visualise

the directions to adjust the proper posture. A yoga expert scored 93.45% of the frames in their assessment of the recognition of posture correctly.

A brand-new dataset called Yoga-82 was put up by Manisha et al. [12] for the fine-grained classification of human yoga poses. Within their proposed Yoga-82 dataset, they modified the DenseNet201 model to categorise yoga postures, and they were successful in achieving the top-1 score for image classification performance at 79% of the 82 multi-classes.

On the public yoga asanas dataset [13], Kumar et al. [1420] proposed yoga posture estimation using the OpenPose algorithm, a keypoint detection method. Six yoga poses—the cobra pose, lotus pose or sitting pose, corpse pose, mountain pose, triangle pose, and tree pose— were used to demonstrate the technique.

The goal of this project was to create a transfer learning-based coaching system for yoga postures. We gathered data on yoga postures for 8 different classifications, and we suggested a coaching system for yoga postures based on transfer learning with pre-trained weights taken from the CNN architecture model to identify poses in real time. Additionally, our yoga posture coaching system offered feedback on posture instructions when a user practised yoga in front of the system.

4. METHOD AND MODULES

The method incorporated can be divided basically into 4 major steps where different modules and libraries of Python can be used. The 4 steps are:

A. Pose extraction
B. Key-point extraction
C. Application of ML Algorithm
D. Error estimation and feedback.

A. Pose extraction:

Python provides various libraries for image and video processing. OpenCV is a vast library that helps in providing various functions for image and video operations. With OpenCV, we can capture a video or a picture from the camera or local system. cv2.VideoCapture(): The function used to access the webcam. cv2.imread(): Function is used to read a particular frame from the image or video. cv2.imshow() is used to show the frames from the webcam. Desired operations on that video or image are performed further.

B. Key-point extraction:

A webcam collects real-time images while performing the yoga pose. The pose extracted from the OpenCV is now used to extract 33 key points, with the help of Media pipes from the body that helps us to classify and determine the errors while the pose is performed. The Media Pipe Pose Land marker task detects the landmarks of human bodies in an image. This helps to identify key body locations and render visual effects on them. This task uses machine learning (ML) models that can work with single images or a continuous stream of images. The output of body pose landmarks in image coordinates and in 3-dimensional (x,y,z) world coordinates. Media pipe also helps us to draw the pose extracted using the util function. The extracted points are then compared to set point values in the training data based on the positions and angles between them. The 33 key points extracted can be shown below:

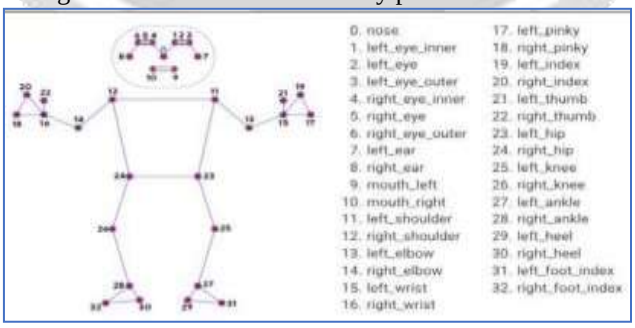

Fig-1: Media pipes pose extraction

C. Application of ML Algorithm:

As the pose can be identified by classifying itself, thus for image classification we are using Convolution Neural Network. A neural network type called a convolutional neural network, or CNN or ConvNet, is particularly adept at handling input with a grid-like architecture, like an image. A binary representation of visual data is a digital image.

It is made up of a grid-like arrangement of pixels, each of which has a pixel value to indicate how bright and what color it should be.

The three layers of a CNN are typically convolutional, pooling, and fully connected.

- Convolution layer:

The foundational component of the CNN is the convolution layer. It carries the majority of the computational load on the network. This layer creates a dot product between two matrices, one of which is the kernel—a collection of learnable parameters—and the other of which is the constrained area of the receptive field.

- Pooling layer:

By calculating an aggregate statistic from the surrounding outputs, the pooling layer substitutes for the network's output at specific locations. This aids in shrinking the representation's spatial size, which lowers the amount of computation and weights needed. Each slice of the representation is subjected to the pooling operation separately.

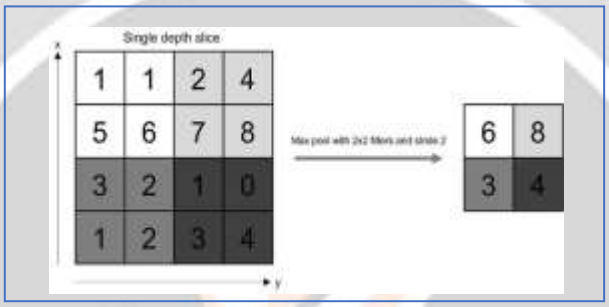

Fig-2: Pooling layer of CNN

- Fully Connected Layer:

As in a regular FCNN, all neurons in this layer are fully connected to all neurons in the layer before it and the layer after it. Because of this, it can be calculated using the standard method of matrix multiplication followed by a bias effect. The mapping of the representation between the input and output is facilitated by the FC layer. The Rectified Linear Unit (ReLU) has gained a lot of popularity recently. The function ()=max (0,) is computed. To put it another way, the activation is just a threshold at zero.

The CNN has an architecture as follows:

[INPUT]
→[CONV 1] → [BATCH NORM] → [ReLU]
→ [POOL 1]
→ [CONV 2] → [BATCH NORM] → [ReLU] → [POOL 2]
→ [FC LAYER] → [RESULT]

The random forest classifier is utilised for real-time webcam classification, producing good accuracy, and the CSV coordinates are learned collectively. A well-known machine learning method from the supervised learning approach is Random Forest. It can be used to address machine learning problems with classification and regression.
A classifier called Random Forest averages several decision trees applied to various subsets of a given dataset to improve the projected accuracy of that dataset.

The two phases of Random Forest's operation are building the random forest by mixing N decision trees and making predictions for each tree that was built during phase one.

The following steps can be used to explain the procedure:
i: Choose K data points at random from the training set.
ii: Create the decision trees linked to the subsets of data that have been chosen.

iii: Decide on N for the decision trees you want to construct.
iv: Repetition of Steps 1 and 2.
v: Find each decision tree's predictions for any new data points, then group the new data points into the category with the most support.
As a result, CNN and Random Forest Classifier correctly identify or categorise the stance.

D. Error Estimation and Feedback:
We can assess the accuracy of the executed pose by providing users with error estimation and feedback based on the trained model. Angles between critical joints are computed using the help of media pipes. The coordinates returned from the training dataset are compared with the calculated angles. If a mismatch is detected, a text message and a voice message are output to guide users in making necessary adjustments to the current pose and eventually correcting the error.

YOGA TRAINER WITH AI

According to the steps mentioned above, the project Yoga Trainer with AI uses the following methodology. Firstly, the users have 2 options to check their pose accuracy and feedback. Firstly, the image or frame is collected from the user through a webcam or through a file/image uploaded by the user with the help of OpenCV. With the help of the media pipes library, a person present in the frame is identified. Next, the identified subject's

x, y, and z coordinates with respect to positioning are extracted. These are stored in a CSV file which helps in training the model further and calculating the angles between the important joints. Now the frame is sent for frame classification which takes place with the help of CNN and random forest classifier algorithm using the techniques mentioned above. The pose performed by the user is identified and the accuracy of pose classification is returned. Based on the angles between the critical joints of the pose performed, the errors are calculated, and the feedback to the user on how well the pose is performed or to correct a particular pose is given as the output to the user in the form of voice or text message is provided as the output.

Algorithm of Yoga Trainer with AI:

Step 1: Start
Step 2: Input from the user through a webcam or image.
Step 3: Input passed to Mediapipe library.
Step 4: 33 Key point extraction from the person identified via the media pipe.
Step 5: If the user has chosen image classification, then with the help of the trained dataset, get the result through the Convolutional Neural Networks
Step 6: If the user has chosen real-time classification via the web, then, classify the pose via the CNN algorithm and the random forest classifier.
Step 7: Based on the trained dataset the accuracy of the pose performed is calculated and errors are detected based on the coordinates obtained.
Step 8: The above algorithm is implemented using a website where the user interactions with the model are more convenient.
Step 9: The user data is stored in the database through which the accuracy and improvements that happened over time is monitored.
Step 10: Stop

5. RESULTS AND ANALYSIS

The model-Yoga trainer with AI was tested through different stages so that the accuracy and functionality of the model increases. The model effectively the users to explore and learn about various yoga poses in this model. The user can perform the poses with live monitoring where the pose performed is sent to the system as an image or as a webcam feed is fed to mediapipe to extract the body coordinates where the coordinates extracted are trained together and classified for pose identification as shown in the figure below. Then, the accuracy of the pose and the errors while performing are monitored and given to the user as an output. The output is stored along with user details so that the users can track their performance and improve.

Fig-3: Yoga pose identification

6. CONCLUSION

Yoga is an easy to perform exercise at home. But just following a video tutorial without a live instructor can be dangerous. Un-correct yoga poses can lead to various health issues like fractures, sprains, and muscle deformations thus, the need for Yoga Trainer with AI is seen. With this project, users are given error estimation and feedback based on the trained model that is CNN and Random Forest Classifier, which helps us determine the pose performed and the accuracy of the pose executed. Angles between critical joints are computed using coordinates retrieved from the trained dataset. If a mismatch is detected, a text message and a voice message are output to guide users in making necessary adjustments to the current pose and eventually correcting the error.

7. REFERENCES

[1] Wearable X. https://www.wearablex.com/. Accessed 16 July 2021

[2] MIRROR.https://www.mirror.com/. Accessed 16 July 2021

[3] SmartMat.https://www.smartmat.com/about/. Accessed 16 July 2021

[4] YOGANOTCH.https://yoganotch.com/. Accessed 16 July 2021

[5] Chiddarwar GG, Ranjane A, Chindhe M, Deodhar R, Gangamwar P (2020) AI-based yoga pose estimation for android application. Int J Inn Scien Res Tech 5:1070–1073

[6] Anusha M, Dubey S, Raju PS, Pasha IA (2019) Real-time yoga activity with assistance of embedded based smart yoga mat. In: 2019 2nd International Conference on Innovations in Electronics, Signal Processing and Communication, https://doi.org/10.1109/IESPC.2019.8902371

[7] Charles J, Pfister T, Magee D, Hogg D, Zisserman A (2016) Personalizing human video pose estimation. arXiv arXiv:1511.06676

[8] Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y (2019) OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 43:172–186. https://doi.org/10.1109/TPAMI.2019.2929257

[9]Zhang Z (2012) Microsoft kinect sensor and its effect. IEEE Multi 19:4–10. https://doi.org/10.1109/MMUL.2012.24

[10] Chun J, Park S, Ji M (2018) 3D human pose estimation from RGB-D images using deep learning method. In: Proceedings of the 2018 International Conference on Sensor, Signal and Image Processing, pp 51–55. https://doi.org/10.1145/3290589.3290591

[11] Chen HT, He YZ, Hsu CC (2018) Computer- assisted yoga training system. Multi Tools App 77:23969–23991. https://doi.org/10.1007/s11042-018-5721-2

[12] Verma M, Kumawat S, Nakashima Y, Raman S (2020) Yoga-82: a new dataset for fine- grained classification of human poses. arXiv arXiv:2004.10362

[13] Manap MS, Sahak R, Zabidi A, Yassin I, Tahir NM (2015) Object detection using depth information from Kinect sensor. In: 2015 IEEE 11th International Colloquium on Signal Processing & Its Applications. pp 160–163. https://doi.org/10.1109/CSPA.2015.7225638

[14] Karbasi M, Bhatti Z, Nooralishahi P, Shah A, Mazloomnezhad SM (2015) Real-time hands detection in depth image by using distance with Kinect camera. Int J Inter Things 4:1–6. https://doi.org/10.5923/c.ijit.201501.01

[15] Chhaihuoy Long, Eunhye Jo & Yunyoung Nam (2021) Development of a yoga posture coaching system using an interactive display based on transfer learning.