Leveraging OpenCV for Precise Yoga Pose Estimation and Reducing Injury Risks

MSc Research Project
Data Analytics

Anand Jha
Student ID: x21157251

School of Computing
National College of Ireland

Supervisor: Teerath Kumar Menghwar

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Anand Jha
Student ID:	x21157251
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Teerath Kumar Menghwar
Submission Due Date:	14/12/2018
Project Title:	Leveraging OpenCV for Precise Yoga Pose Estimation and Reducing Injury Risks
Word Count:	7105
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Anand Jha
Date:	14/12/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☑
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☑
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☑

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Leveraging OpenCV for Precise Yoga Pose Estimation and Reducing Injury Risks

Anand Jha
x21157251

Abstract

This project aligns yoga practices with AI forefront technology to demonstrate a real-time yoga pose detection system. Using a diverse training set and MediaPipe's Blazepose model, landmark points were extracted and organized for machine learning. The top-performing model was – XGBoost, which demonstrates better accuracy with the complex spatial relationships observed among different yoga poses. Inclusive and personalized guidance is provided by the system using text-to-speech, visual cues and pose-specific benefits. It is this XGBoost's 92% accuracy that underlies its performance in seamless pose classification. The system provides real-time feedback about pose correctness, total pose duration and helpful insights. These features include pose analysis, visual overlays, and auditory guidance to meet the needs of different users. This blend of AI-powered pose detection, feedback loop mechanism and inclusive design defines the shift in paradigm towards opening up a yoga experience for all practitioners.

1 Introduction

Practised for over two millennia to date, beginning in 1500 BC, TYoga revolves around the three-fold human-environment union. Yoga, the term being derived from 'Yuj,' a Sanskrit word for 'union', intends to harmonise individual consciousness and universal tapestry. Yoga is an ancient science that has been valued for thousands of years and occupies an important place in the culture and values of India, where it was born and transformed into something spiritual. In the contemporary world, more and more people are coming to realise the value of personal well-being. Thus, yoga is now accepted as a health-giving practise in parts such as physical exercise and mental therapy. The appeal of this practise goes beyond the development of physical strength to instilling mental serenity and emotional equilibrium. Yoga is deeply anchored in Indian cultural heritage and has an estimated 300 million followers. Practised by people of all ages and societal backgrounds, it easily transcends any limitations or boundaries. The validity of the combination of ancient wisdom and modern aspirations may be especially identified in today's world, as every person tries to have a complex health approach nowadays. That is what the beauty of yoga is—its simplicity. The fusion of yoga with artificial intelligence (AI) as such becomes a pioneering gate to democratise this ancient wisdom by embracing the synergy between tradition and innovation. All it calls for is a space to practise, hence making it accessible to all. Convergent aimed at developing an easy-to-use platform using technology to enable yoga to be relevant in everyone's life by being transformative.

Many research papers Kotak et al. (2022); Mitrović and Milošević (2022) have been written based on simply using math angles. Some researchers Min (2022); Nagarkoti et al. (2019) combine computer vision and deep learning methods to identify the human pose in the video, whereasMroz et al. (2021) have done a comparison study between the two famous Open Pose and Blazepose landmark detection models. In the papers by Mili et al. (2022) research is done to identify the pose of the cricket game players. Many researchers Palanimeera and Ponmozhi (2021); Dedhia et al. (2023); Munea et al. (2020) have dedicated their work on yoga pose detection with the help of machine learning techniques. This project aims to integrate the ultimate wisdom of yoga with AI's cutting-edge technologies. This will make the transformative power of yoga easily accessible to anyone, irrespective of their background or physical location. The primary goal is to preserve yoga's legacy, enhance its value as a public good, and disseminate it in such a way that all people use it as a means of total welfare.

1.1 Motivation and project Background

Blending technology with yoga offers a path to affordable, customizable practices in a new era of holistic well-being. The source of this project's motivation is the use of AI in enhancing yoga sessions providing immediate feedback and involving practitioners of different people with ethnic abilities. Inspired by increased interest in the physical and mental wellness benefits of yoga and enabled by research advancements in computer vision and AI methods, this project aims to fulfil the need for precise pose detection, individualized guidance, and instantaneous feedback. This undertaking intends to fuse traditional yoga with new technological trends for the benefit of every learner from the worlds over 300 million practitioners.

1.2 Research Question

RQ: *"How can computer vision and machine learning be effectively combined to develop a real-time yoga pose detection system that provides accurate pose classification, guided feedback, and pose benefit analysis to enhance user experience and accessibility for yoga practitioners?"*

Sub RQ:*" How can the identified yoga poses, and their corresponding benefits be integrated into the system to provide comprehensive guidance and enhance user engagement during yoga sessions?"*

To solve the research question, the following objectives are specified and implemented:.

1.3 Research Objectives

Objective	Description
Obj 1	A Critical review of literature on human exercise pose detection (2020-2023)
Obj 2	Develop a robust data gathering module to capture real-time yoga poses via webcam or external cameras, incorporating OpenCV for image processing.
Obj 3	Implement a pose detection module utilizing MediaPipe's Blazepose to extract 33 landmark points in 3D space (X, Y, Z) representing yoga poses.
Obj 4	Integrate a pre-trained machine learning model, to classify detected poses based on extracted landmark data.
Obj 5	Create a feedback module to provide real-time guidance and benefits analysis based on pose accuracy, incorporating both audio and visual cues.

Table 1: Objectives

The rest of the technical report is structured as follows: chapter 2 presents an investigation of previous literature in exercise pose detection using OpenCV and machine learning algorithms; based on the literature, chapter 3 presents methodology; and chapter 4 focuses on implementation and evaluation.

2 Related Work

Exercise pose analysis may be complicated, and numerous researchers have experimented with various methodologies in an effort to develop better-performing models that can be used to estimate the pose of the users. Exercise analysis can be complicated and provide numerous difficulties. The analysis of numerous studies that attempted to resolve this problem by utilising various methodologies, algorithms, etc. is presented here in the literature review. Depending on the chosen model, which is detailed below, the selection is separated into 3 sections.

2.1 Pose Detection Using Joint Angle

A groundbreaking revolution is taking place in the fitness industry with the integration of AI and computer vision. Researchers Faujdar et al. (2023) have explored the potential of using AI to track gym trainers, analyse workout repetitions, and estimate body poses to understand athlete movements. They discuss both bottom-up and top-down approaches to computer vision methodologies. The use of AI has led to workout suggestions, the rising popularity of AI trainers, and the adoption of wearables in fitness. The researchers emphasise the importance of posture estimation achieved through neural networks and convolutional neural networks. This paper primarily focuses on pose estimation techniques, interpretation of coordinates, and utilising MediaPipe and OpenCV for gym tracking. Overall, it provides a perspective on how AI and computer vision are transforming the fitness industry, specifically in human pose estimation and tracking, with room for further advancements.
Another research paper Mitrović and Milošević (2022); Mroz et al. (2021) introduces a machine learning model aimed at preventing injuries during activities by utilising MediaPipe to track body poses. This system enables professionals to identify angles accurately and

study body postures for injury prevention purposes, and analyse physical activity patterns. The system is designed to analyse video inputs and angles, offering information about body alignment and joint angles. It can quickly estimate posture and angles by examining videos of people engaging in activities, resulting in a comprehensive set of results. Possible future developments include making it web-based for accessibility, processing videos in real time, and potentially integrating it with incorrect exercise detection for complete guidance without the need for supervision. Altogether, this study presents an integrated approach that combines methods for measuring posture with implications for computer-assisted posture estimators and reducing injuries during physical activities.
This research paper Dsouza et al. (2020) examines human pose estimation in computer vision in the context of an AI-powered gym trainer. including the complexity of body articulation and appearance variability, which have become more prevalent in recent years due to usage of RGBD cameras for depth-based estimation. In this paper, we propose a machine learning-based approach for accurate pose estimation in which professional workout data is used to train the system and a GUI is provided for interaction. For body part identification and segmentation, it uses tools such as OpenCV, convolutional neural networks (CNNs) and models such as COCO (MPII). The development of the system encompasses some stages, starting with model acquisition, virtual skeletal structure generation, determining joint angles, and visual comparisons of user movements and exercises done. The visual results presented are stick figure diagrams, angles of the joints, and graphs. In future, we have set goals of enhancing accuracy by utilising high-quality cameras and biomedical uses. In conclusion, this paper deeply addresses the hindrances and methods of having an AI-powered gym trainer using human pose estimation with OpenCV.

2.2 Deep Learning Algorithm

Pose estimation is important in human action recognition in the article Gadhiya and Kalani (2021), which discusses deep learning, especially with regards to Convolutional Neural Networks (CNN) in increasing accuracy. It describes push-up posture detection model and reviews models such as OpenPose, BlazePose, DeepPose and Hourglass architectures for different pose estimations. It moderately explores possible benefits of pose tracking over traditional systems, focusing on the first specific usage area – predicting fine-grained details (e.g., pedestrian). It classifies pose estimation approaches into bottom-up and top-down approaches relying on kinematic models. The ins and outs of OpenPose's Part Affinity Fields, BlazePose's mobile-friendly design, DeepPose's regression technique, and Hourglass' encoder-decoder network are analysed. 2d push-up posture detection: Performance evaluation of openpose and blaze-pose for model selection in different applications. It highlights the importance of CNNs and provides a detailed review of current pose estimation models, their utilities, as well as evaluation criterion. Similarly Bazarevsky et al. (2020) has done research on CNN deep learning model.
This paper Liaqat et al. (2021); Mohammed et al. (2022) proposes a novel posture detection technique using an integration of machine learning (ML) and deep learning (DL) classifiers that recognises standing, sitting, and walking postures for continuous monitoring of elderly people's health. This involved a hybrid model of ML (e.g. support vector machine, decision tree), and DL (e.g. 1D/2D convolutional neural networks, Long short-term memory) due to inadequate traditional ML classifiers. Posture prediction employed feature extraction, including skew, percentile, square root among others. Out

of all the different classifiers, the hybrid DL approach performed very well to an accuracy of 98 percent, surpassing traditional ML and DL approaches. This indicates that the performance of this method was better that existing ones for a certain problem, posture detection. The study's strong methodology and comparative analysis highlight the hybrid ML-DL model's prospects for remote health monitoring in vulnerable populations, providing key guidance for further research in human sensing and artificial intelligence.
In this paper Chen et al. (2023), a new CNN-LSTM model for Baduanjin action recognition has been presented in order to tackle limitations of current geometric-like action recognition models. The model presented in this paper performs better by leveraging both Convolutional Neural Networks (CNNs) and Long short-term memory (LSTM) components and achieves 96.43% accuracy on the test set compared to 66.07% of accuracy obtained with the former model. (Convolutional Neural Networks) CNN's ability to retrieve features of an image contributes much to the accuracy of classification by Long short-term memory (LSTM). The model presents potential as an intelligent rehabilitation assessment for home-based practise and proves useful for distinguishing intricate actions, filling the literature void on traditional Chinese exercises.The CNN-LSTM network takes advantage of (Long short-term memory) LSTM's ability to recognise sequential processes and CNN's capability to extract features from video frames. This paper documents an elaborate experimental setup, evaluation metrics, and results to summarise the effectiveness of the model over manual feature-based techniques. The research points to possible applications in smart rehabilitation, health monitoring, and sport skill evaluation. Such model sensitivity and adaptability will be further improved in future developments, with the goal to instance it into exercises involving fast movements.
In the research article Rahman et al. (2022), a 3-Layer CNN-LSTM model is proposed for automating the detection and classification of rehabilitation exercises performed by stroke patients during physiotherapy. The RGB camera data was collected in the study, preprocessed, and Convolution neural networks (CNN) used to extract features, accompanied by FC layers and an Long short-term memory (LSTM) layer for temporal dynamics learning. The model apparently emerges excellently compared to other algorithms, with an accuracy level of 91.3%. The article highlights the first cost of stroke, presents new developments in virtual reality rehabilitation, and describes state-of-the-art methods in human activity recognition. It provides an in-depth discussion on Convolution neural networks (CNN) and Long short-term memory (LSTM) algorithms, data pre-processing, the model proposed, implementation discussed, performance evaluation addressed, and finally the experimental results given.

2.3 Machine Learning Algorithm

The research paper Agrawal et al. (2020) deals with AI software development for Yoga Practice Recommendation, where accuracy is improved by exploiting pose detection using the tf-pose algorithm. tf-pose is used to generate joint coordinates and angles, which are utilized as features for the machine learning models in a dataset consisting of 5500 images of ten yoga poses. These models include Logistic Regression, Random Forest, SVM, Decision Tree, Naïve Bayes and K-Nearest Neighbours (KNN), but Random Forest has been used with an accuracy of 99.04%. The creation of YOGI dataset was a result of challenges in the accuracy of the YOGA dataset that contained different poses with difficult leg and hand folds. The current study investigates such techniques as Kinect, tf-pose, CNN, and LSTM useful in high-accuracy yoga pose detection. Angles from tf-pose's skeleton joints

give a total of 94.28% accuracy on machine learning models. We highlight the importance of Random Forest and other models in assessing the accuracy of AI software for yoga practice assistant. It underscores the dataset precision issues and various techniques such as tf-pose and CNN, used to improve yoga pose detection and correction.
The research by Shah et al. (2021) presents an AI supported system for yoga pose recognition and correction achieved through various machine learning techniques including pose estimation, K-Nearest Neighbors (KNN) classification, Posenet and a K-Nearest Neighbors (KNN) classifier model. The major goal of the system consists in real-time recognition and correction of yoga poses that utilise a K-Nearest Neighbors (KNN) classifier model, which is remarkable due to its 94.4% correctness rate. The joint localization and complex joint relations modeling utilizes advanced techniques such as OpenPose and Convolutional Neural Networks (CNNs). The paper underscores the importance of machine learning in developing stable systems for detecting yoga poses and presents future research opportunities such as utilisation of advanced image categorization methods, enlarging pose databases, and customising yoga poses according to individual characteristics thereby providing an additional scope for more advanced ML based research and enhancements of the existing system.
Pose Estimation and Machine Learning for Identifying Yoga Poses Bahukhandi and Gupta (2021). It adopts open-source data comprising videos that contain the performance of six yoga poses by volunteers. It uses a combination of bottom-up and top-down procedures with the Mediapipe method to estimate poses in real-time videos. This study addresses machine learning classifying, whereby it highlights algorithms such as logistic regression and support vector machines. It details the system's phases: experimental setting; data collection, preprocessing, and feature drawing. This includes explaination of the dataset overview, data processing, normalization and feature engineering via joint angle calculation. The research applies classification-based algorithms to train and evaluate models, where logistic regression achieves an accuracy rate of 94%. This study developed a yoga pose classifier through pose estimation and machine learning, highlighting emphasising setup through model evaluation stages that involved a range of algorithms and methods.
This research paper Palanimeera and Ponmozhi (2021) discusses the use of machine learning algorithms in relating to accurate classification of yoga poses with main emphasis on Sun salutation postures. This study tries to identify these poses while they are developed using machine learning models and pose estimation algorithms. These authors employed different machine learning techniques such as K-Nearest Neighbors (KNN), Support vectir Machine (SVM), Naïve Bayes, and logistic regression with the best accuracy of 96% using KNN algorithm. The paper presents various problems in human pose estimation with a focus on accurate posture detection and its importance in musculoskeletal disorders, video surveillance, biometrics, etc. It also shows how pose recognition has grown from 2D to 3D and the improvement of pose estimation algorithms, including generative and discriminative methods. The outcomes of the study show great accuracy in classifying yoga poses and serve to demonstrate that machine learning is important in solving musculoskeletal problems as well as enhancing better posture recognition in a plethora of domains.

3 Methodology

In the process of creating a robust real-time yoga pose detection system, this research followed a structured research methodology, amalgamating CRISP principles and a distinct five-step process: Data Gathering, Data Preprocessing, Data Transformation, Data Modelling, and Evaluation as shown in Figure 1.

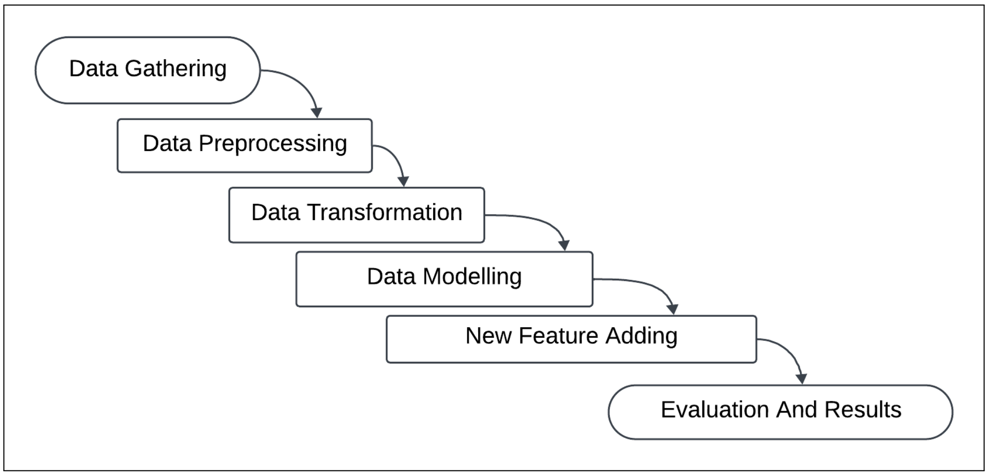

Figure 1: Methodology Flow Chart

3.1 Data Gathering

It was commenced by assembling a dataset of popular yoga poses found in Kaggle dataset [1], for each pose 1304 images were collected. There was a separate folder for each of the poses by the corresponding name. Yoga Asana pose was named "ADHO MUKHA SVANASANA", "VRIKSHASANA", "BALASANA"," UTKATA KONASANA", VIRABHADRASANA". Figure 2 shows total number of images of each exercise.

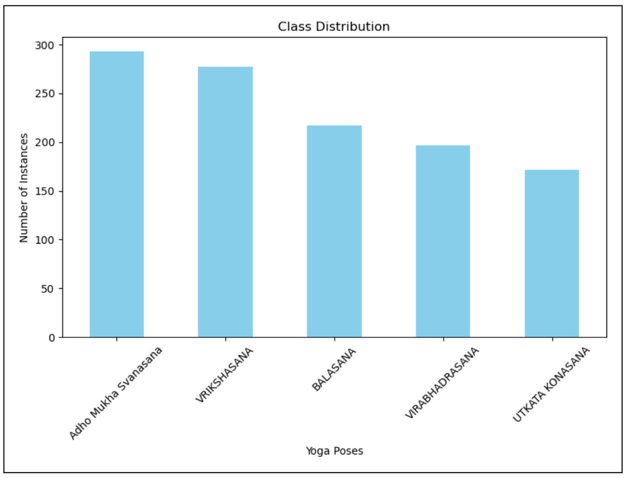

Figure 2: Class Distribution

[1]Dataset Link: `https://www.kaggle.com/datasets/swapzzj/yoga-dataset?select=Dataset`

3.2 Data Processing

Using MediaPipe's Blazepose model along with OpenCV, we convert this yoga pose images into skeleton images and extracted from each image 33 3D landmark points as shown in the Figure 3 (x, y and z coordinates), so there will be total 99 features column. The ability of Blazepose to accurately identify key points has made it relevant in fitness application and real-time CPU inference. The were appended to a CSV file along with their corresponding class name for the yoga pose.

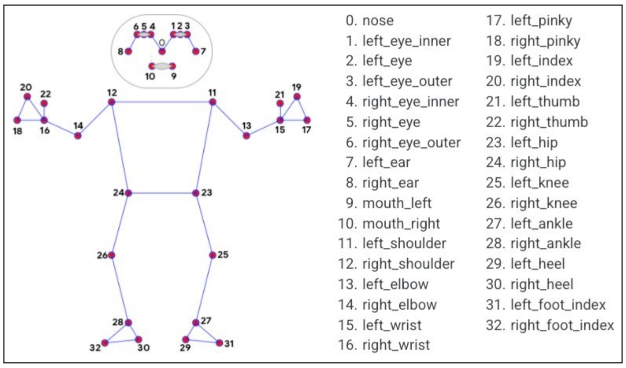

Figure 3: BlazePose Model - Topology as depicted in the original paper Bazarevsky et al. (2020)

3.3 Data Transformation

At this stage, the landmark points were standardized, and the dataset was divided into dependent and independent variables. The response variable indicating yoga poses was label-encoded, and the dataset was divided into training and testing sets in a 70: 30 ratios.

3.4 Data Modelling

After this the various machine learning (Random Forest, XGBoost, Support Vector Machine (SVM), Decision Tree) and deep learning (Long short-term memory (LSTM), 1D CNN(Convolutional neural networks) models were trained on and evaluated using the same training dataset by the author. This phase included hyperparameter optimization, cross-validation, reshaping data for deep-learning models, and model training. Categorical Accuracy, Loss, Precision and Recall defines diverse evaluation on models.

3.5 New Feature Added: Enhanced User Experience and Accessibility:

The yoga-benefit function integration introduces live pose-specific benefits that adds value to the user experience through enlightened insights on each recognized pose. In addition, the incorporation of text-to-voice capabilities and graphical cues enables ease of use for different users including visually impaired thereby promoting an augmented interaction model.

3.6 Evaluation and Results:

All the models were evaluated based on how accurate they were, as well as, their precision, recall and time complexity. The Confusion metrics, the confusion reports as well as the accuracy/loss plots were analyzed. After evaluation of the models, XGBoost model became their best choice for real-time yoga pose prediction leading to development of real time yoga pose detection with Blazepose, XGBoost Classifier and computer vision where users received instant feedback. In this proof of concept, we developed a real-time yoga pose detection system using state of the art technology in combination with systematic analysis, which forms a structured approach offering instant feedback during yoga practice. Integration of advanced techniques and rigorous evaluation methods makes the system accurate and reliable for recognizing yoga poses and their guidance to the people practising them.

4 Design Specification

In order to build the robust real-time exercise pose detection system, it is divided into four different parts, as shown in Figure 4.

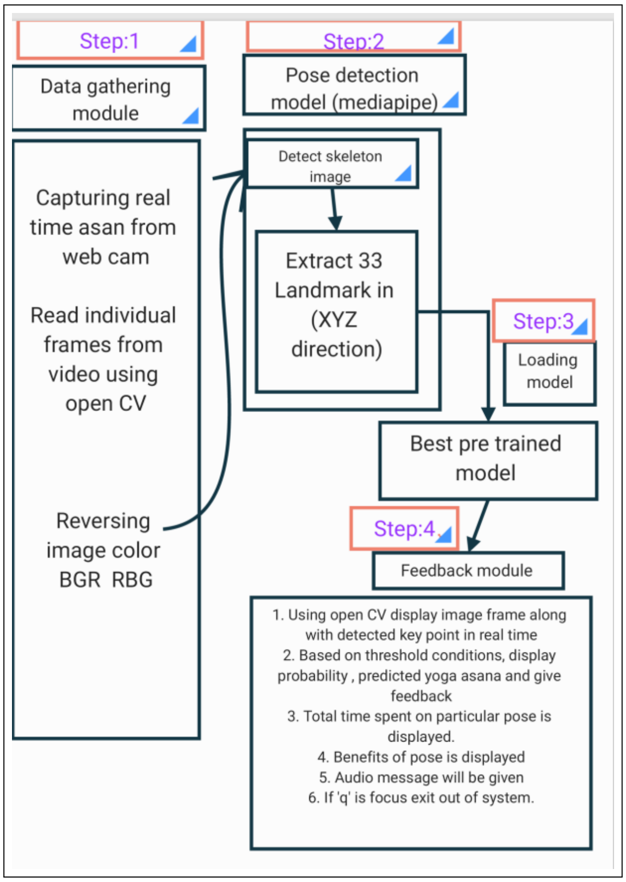

Figure 4: System Architecture

4.1 Data Gathering Module

The system initiates by capturing real-time ASANA videos from the user's webcam or an external camera. Each frame of the video undergoes processing using OpenCV, where image scaling and conversion from BGR to RGB format occurs to align with the requirements of subsequent modules.

4.2 Pose Detection Module

The processed frames are then forwarded to the Pose Detection Module. Here, MediaPipe, an integral part of the system, is employed to identify the skeleton image within each frame. This process involves the extraction of 33 landmarks in 3D (X, Y, Z) directions. These landmarks represent key joint points, with X and Y denoting distances from the origin and Z representing the distance from the camera.

4.3 Loading Model Module

The extracted 33 landmarks are fed into the system's pre-trained model, the XGBoost Classifier. This model has been fine-tuned and optimized for accurately classifying yoga poses based on landmark data. This module handles the interpretation of landmark configurations and their association with specific yoga poses.

4.4 Feedback Module

Upon classification, the system enters the Feedback Module. Here, the detected frame's 33-keypoint skeleton is displayed. The system applies threshold-based filtering on the classification probability. If the probability surpasses 0.85, the detected yoga asana is visually displayed along with an enthusiastic "Very Good" feedback message, complemented by an audio message. Additionally, the associated benefits of the yoga pose are presented. For probabilities between 0.80 to 0.85, the system provides feedback indicating the detected asana and messages "Good" to encourage improvement. When the probability falls below 0.80, the system prompts for improvement and displays the total time spent in that yoga pose across all conditions. The user can exit the system by pressing the 'q' key on the keyboard. This architecture systematically processes real-time ASANA videos, employing advanced techniques for pose detection, model interpretation, and user feedback, enhancing the user's yoga experience while providing tailored guidance and encouragement.

5 Implementation

For this experiment, a set of yoga poses' dataset was collected, which were used as diverse samples; in addition to this, MediaPipe's Blazepose model was acquired to extract out landmark points that correspond with body alignment. On the other hand, detect yoga poses based on just joint angles Faujdar et al. (2023) faced some difficulties which were attributed to the similarly posed used thus making them important since they are confusion and misclassification of yoga pose Liaqat et al. (2021). Therefore, the study transformed to machine learning models such as XGBoost; Convolutional neural networks (CNN);

Long short-term memory (LSTM); K-nearest neighbors (KNN); decision trees and Support vector machine (SVM) by training different algorithms using preprocessed landmark data. XGBoost model was found to be the most accurate in classifying yoga pose. To detect the human pose in real-time by integrating OpenCV and MediaPipe, an XGBoost model was established to precisely recognize yoga poses in images. Other functionalities include pose benefit presentations, vision implementation and runtime performance improvement based on user feedback. Robustness of the system was achieved by validation and testing across numerous scenarios, making comprehensive. These elements contributed to the success of the system in accurate classification, inclusive design and overall improvement for all users in yoga.

6 Evaluation

A comprehensive evaluation was performed to review the results of the experiments which included analysis on system performance, user-oriented design and position in comparison with other AI-based yoga pose detection systems. The evaluation examined the accuracy, user engagement levels, and positioning of the system in light of other research works to allow readers grasp its strength while also identifying areas where it could be improved. Each experiment was subject to analysis not only to determine its contributions, limitations but also the implications that it holds in advancing the system.

6.1 Experiment 1: Keypoints Detection Using MediaPipe

The project began with MediaPipe's Blazepose model to set the groundwork for a reasonable yoga pose detection. A comprehensive dataset containing various yoga poses was constructed in this trial. The dataset is prepared, which consist of the total 1304 images of 5 different yogasana, collected from the Kaggle dataset. Based on the Blazepose, crucial landmark points transformed to the joints and body parts were extracted from the images of this dataset. These landmark points, 33 in number in 3D co-ordinates (x,y,z direction) formed pivotal references that identified the spatial disposition of different yoga poses. z is the distance from the webcam, while X, Y coordinates refer to the distance from the origin. Figure 5 shows the Input image and output image in which all the 33 landmarks are drawn.

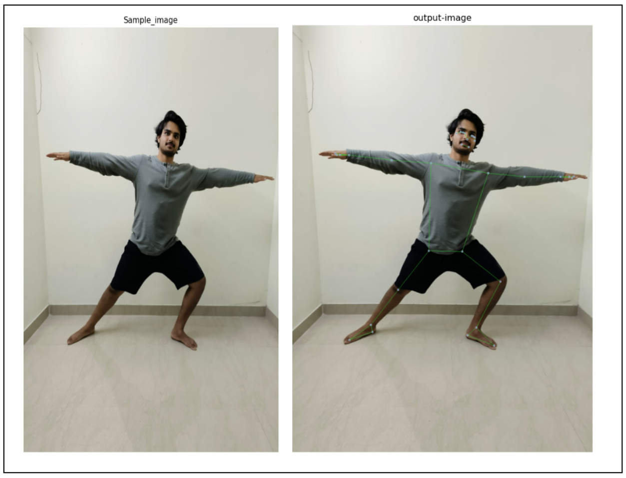

Figure 5: Landmark Detection Using Mediapipe

6.2 Experiment 2: Key Point Detection Using Real-time Video

The project moved from static images to processing the real time video feed. With this model, the system was able to utilize Blazepose to process live video frames thus enabling the dynamic extraction of landmark points from moving images. This photo land- mark real time extraction enabled pose recognition in live video streams as shown in Figure 6. In Figure 6 user is not performing any Yoga pose, therefore system is displaying message "No pose detected, Please let full body come into frame".

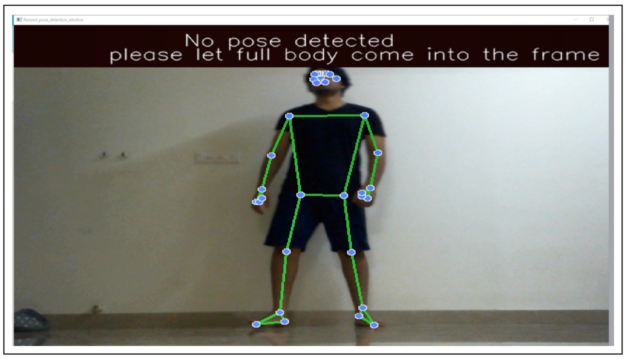

Figure 6: Landmark Detection in Live Video Streming

6.3 Experiment 3: Predicting Approach Using Heuristic Method

The first approach utilized a heuristic method using formulas of joint angles for pose recognition. The methodology revolved around estimations for joint angles, which provided the yoga stances difference. Yet, the difficulties that arose showed that it is hard to interpret poses correctly on the ground of joint angles only. Some of the drawbacks were increased code complexity, specific joint angle ranges requirement that could be problematic for users with less flexibility and confusion where different poses held similar joint angle conditions. Therefore, this experiment emphasized the necessity to use more improved methods.

6.4 Experiment 4: Applying Deep Learning Models (CNN and LSTM)

In the experiment 4, the advanced deep learning such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models were employed to apply more sophisticated techniques for yoga pose classification. The Long Short-Term Memory (LSTM) model that deals with sequential data and temporal dependencies used a layer configuration peculiar to yoga poses categorization. The model involved three LSTM layers with 64, 128, and 64 units for simplifying the nonlinear relations between landmarks. Dropout regularization was carefully set at 20% across all LSTM layers, to restrain the tendency of overfitting. The inclusion of dropout layers in strategic positions interrupted the learning process allowing for generalization of consecutive yoga posture data patterns and linkages. The model had two dense layers with 64 and 32 neurons after the LSTM layers. Such highly connected layers facilitated the extraction of higher-level features and abstraction from the landmark data temporal representations. During training and evaluation, the LSTM model provided an accuracy of 90% on test dataset and 0.41 loss (can be seen in Figure 7) that accomplished all categorization in the model through a five-neuron softmax-activated output layer associated with yoga positions. The exceptional accuracy exhibited portrayed the model's ability to effectively represent and interpret the intricate temporal dependencies in sequential yoga postures, hence validating its pose classification accuracy in this project.

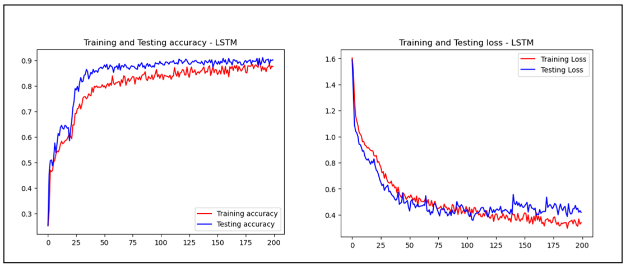

Figure 7: Accuracy and Loss for LSTM

This project developed an accurate Convolutional Neural Network (CNN) model with specific layer structures for recognizing and classifying yoga poses based on landmark data. The architecture began with Conv1D layers comprising of 64 filters and kernel size of 3 in order to extract spatial information from input data. Strategic dropout layers at 0.5 were added after the Conv1D layer and MaxPooling layer that helped in overfitting prevention as Convolutional layers detected intricate patterns in landmark representations. A MaxPooling 1D layer with pool size 1 was added after the convolutional and dropout layers and it optimized the model's ability to learn varied yoga position data attributes by selecting removing neurons during training. After the dropout layers, CNN structure had two thick layers after which were 64 and 8 neurons that down sampled while retaining important information crucial in extraction of key features and patterns yoga position spatial representations. These layers were densely connected and helped the model learn high-level representations and abstract properties from the retrieved landmark data, enabling understanding of yoga poses' complex spatial interactions. As will be discussed later, activation functions were pivotal network wide. In the hidden layers, the model learned non-linear data associations based on the Rectified Linear Unit (ReLU) activation function. The final model of our CNN used 29 convolutions, 3 × 3 filters with the rectified linear unit (ReLU) activation function, and two dropout hidden layers with dropout rates of 10% after each convolution layer. This model had an output layer of five neurons activated through softmax for classification purposes. The performance result using this model was good; after training and evaluation, it achieved an accuracy rate of as high as 91% loss in test on the dataset which means loss is around only to be around 0.38 (can be seen in Figure 8). In this performance, the model that was above to interpret complex spatial patterns and relationships in landmark data, attesting to its yoga position classification precision.

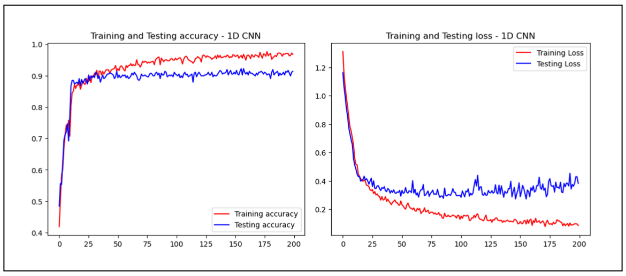

Figure 8: Accuracy and Loss Graph for CNN

6.5 Experiment 5: Machine Learning Algorithm Application

Poses classifications were evaluated with several machine learning algorithms of Random Forest, SVM, Decision Tree, KNN and XGBoost. Out of these, XGBoost had the highest accuracy rate at 92%. The experiment underlined the importance of model selection for best accuracy in pose classification. Table 2 is comparison of evaluation metrics all algorithm used in this project.

Algorithms	Accuracy	Precision	Recall	F1 Score
Random Forest	90.78%	91.07%	89.83%	90.22%
Support Vector Machine	90.78%	91.78%	89.63%	90.01%
Decision Tree	86.46%	85.64%	85.81%	85.71%
KNN	89.34%	82.64%	85.61%	87.54%
XGBoost	91.64%	91.64%	90.65%	91.00%

Table 2: Machine Learning Algorithms Comparision

6.6 Experiment 6: Development of 'detect pose' Function

The team used the 'detect pose' function which was an important development that provided better detection of yoga poses in images. Utilizing sophisticated pose detection algorithms and incorporating them into a simple function simplified the process of pose recognition making it efficient and accurate. Figure 9 shows the result/output of the function detect. In which user is performing "Vriskhasana" pose and then system detect and draw 3d pose (as shown in Figure 9) based on the keypoint detected by the model and draw and predict the pose of the image with the probability of 1.0. In the Figure 9 system first take image from the user performing yoga, it will first take image and feed to mediapipe and find the keypoint and draw the landmark and will also daw the 3d image with the help from the landmark, here importance of Z coordinate can be seen, it uses to draw how far is user away from the camera.

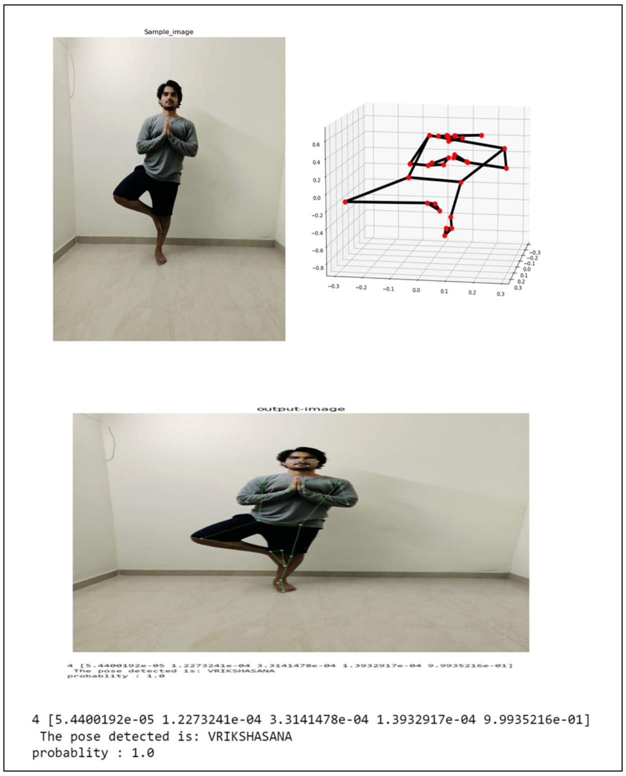

Figure 9: Yoga Pose Image Detection

6.7 Experiment 7: Real-time Yoga Asana Detection and User Interaction

Finally, the system was implemented in real-time to perform yoga asanas detection from live webcam videos. Some of the system features included total pose duration, audio feedback for vision-impaired users, live display of pose benefits and dynamic visual and audio feedback based on pose confidence levels. This led to significant enhancements in the user engagement and accessibility during practice or training yoga. Below are the video screenshots taken of all the exercise performed in front of webcam.

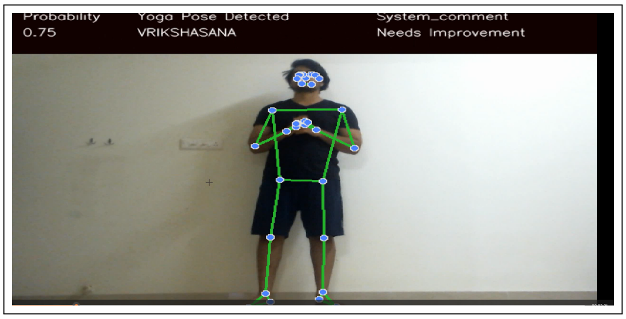

Figure 10: Vrikshasana Pose (Probability 0.75

In Figure 10 it can be seen the user is not performing the yoga pose properly, but the probability of the pose is 0.75 that it is "Vrikshasana" and then it gives feedback that pose "need improvement". Therefore, system will not start the time duration of user in that pose and will not display the benefit of the "Vrikshasana". As user improve the yoga posture and probability is above 0.80, the time duration will be started, and benefit of the pose will be displayed on the screen as shown in the Figure 11. And for the visually impaired user the audio message will be helpful, and, in this way, it helps the user to understand they are doing correct exercise, and it will decrease the rate of injury.

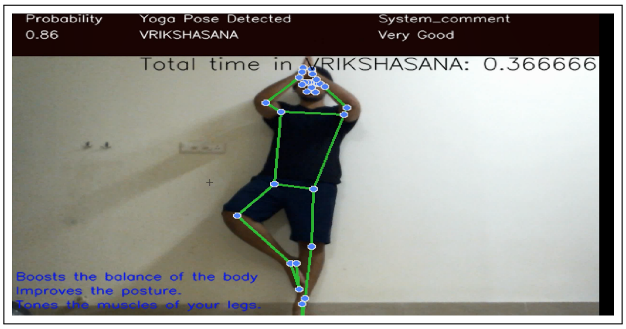

Figure 11: Vrikshasana Pose (Probability 0.86)

https://www.overleaf.com/project/6584a0794fe57e24f8a16524

Figure 12: Virbhadrasana Pose Detection

In Figure 12 user is performing "Virbhadrasana" where the probability of pose is 0.88, therefore system displayed feedback "Very Good" and started the timer and displayed the benefit of doing "Virbhadrasana".

Figure 13: Adho Mukha Svanasana Pose Detection

In the Figure 13 user is performing “Adho Mukha Svanasana” with the probability of 0.95, and system displayed the benefit and feedback of the that pose.

Figure 14: Utkata Konasana Pose Detection

In the Figure 14 user is performing “Utkata Konasana”, its probability was between 0.80 to 0.85, therefore the system display feedback “Good”, and also displayed benefit of performing “Utkata Konasana”. Each experiment played a crucial role in enhancing the system’s accuracy, robustness, and user engagement, addressing specific challenges and refining methodologies for real-time yoga pose detection.

6.8 Discussion

This project consisted of experiments that yielded results. An important success was that the system was capable of correctly identifying yoga poses with an accuracy of 92%, exceeding the achievements made by other AI based systems. The success relies on the use of learning models and choosing the most appropriate machine learning algorithm. Also, it runs on-time and provides a feed-back on the pose completion by means of live webcam videos. There results can be seen in the Figure 10,11,12,13,14. Audio and visual cues are provided that give meaning to the motion sequence. The refinement of pose accuracy is done through these cues. Minimizing the risk of injuries. Furthermore, the system provides response enabling users with disabilities hence makes this product available to a large number of people. This project employs an array of learning models including CNNs and LSTMs for high precision and strong pose detect ability. Finally, an effective “detect pose” function that is optimized for yoga pose recognition from images has been created and can be seen in Figure 7 that it can predict the pose image accurately. Unlike others, it also gives actual information in real time which helps a user to shape their behaviour accordingly and minimizes chances of injuries. Further, there is an additional type of audio feedback where users that have a vision impairment will benefit more from the high levels of accessibility. In general, this project is an addition to the field of AI-powered yoga pose recognition in general. With these features and characteristics, it becomes a priceless support for yogi practitioners as to its impressive accuracy, dependability, and user-friendliness. Hence it answers both research question and research sub question by completing all the research objective mentioned in chapter 1.

7 Conclusion and Future Work

The transformative system is an example of our work that merges yoga with advanced computer vision and machine learning. This system is more than a pose detector and assists yoga enthusiasts to practice poses correctly and improve over time by offering real-time feedback and informing the benefits related to yoga asana. XGBoost ranked highly for classifying yoga poses among all the machine learning used in this project, which constituted the most accurate and reliable model in this field. It answers both research questions and research sub questions by completing all the research objective mentioned in chapter 1. To summarise this project created a yoga pose detection that can detect the pose but also have feature to display the name of the pose, duration spent in the pose, probability of the pose feedback of the pose based on probability and by adding the audio feature to the system, it will not only help the normal people but also to the visually impaired users. Although it can be said that it is perfect system to detect the yoga pose, there are some areas on which future work can be done. This project considered only 5 yoga poses to build the model, but in future more yoga pose dataset can be taken to create more robust model. The number of the sample data can be increased, or image augmentation can be done. Apart from the machine learning algorithm used in this project future researchers can use different models to check whether it increases the accuracy of the system. GNN (Graphical Neural Network) can also be used because it considers the body as graph and joint as node and it considered as best for image classification. Researchers can also take help of yoga instructor to create own and accurate data to train the model. It can also be integrated with the chatbot so users can ask yoga related questions, like which yoga is useful to increase flexibility or increases the abdominal strength. And in yoga along with yoga posture breathing techniques also matters, so new features like while performing the yoga the system can assist the user when to inhale and when to exhale, so that its yoga's full benefit can be taken by the user who is performing the yoga asanas. Similarly, like yoga body building exercise pose estimation can also be done. Where user can do Pull-ups, sit-ups, plank. And new features like repetition counter can also pe embedded into the project.

Acknowlegment

I express my sincere gratitude to my supervisor,Teerath Kumar Menghwar for his invaluable guidance and supervision throughout the duration of this project. His expertise and support significantly contributed to the successful completion of this research endeavour. I would also like to extend my thanks to my friends at college who provided encouragement, shared insights, and engaged in stimulating discussions, enriching the overall experience. Finally, I am deeply appreciative of my family for their unwavering support, understanding, and encouragement during this project. Their constant motivation and belief in my abilities have been a driving force behind the completion of this work.

References

Agrawal, Y., Shah, Y. and Sharma, A. (2020). Implementation of machine learning technique for identification of yoga poses, *2020 IEEE 9th international conference on communication systems and network technologies (CSNT)*, Ieee, pp. 40–43.

Bahukhandi, U. and Gupta, S. (2021). Yoga pose detection and classification using machine learning techniques, *Int Res J Mod Eng Technol Sci* **3**(12): 13–15.

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F. and Grundmann, M. (2020). Blazepose: On-device real-time body pose tracking, *arXiv preprint arXiv:2006.10204* .

Chen, J., Wang, J., Yuan, Q. and Yang, Z. (2023). Cnn-lstm model for recognizing video-recorded actions performed in a traditional chinese exercise, *IEEE Journal of Translational Engineering in Health and Medicine* .

Dedhia, U., Bhoir, P., Ranka, P. and Kanani, P. (2023). Pose estimation and virtual gym assistant using mediapipe and machine learning, *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, pp. 1–7.

Dsouza, G., Maurya, D. and Patel, A. (2020). Smart gym trainer using human pose estimation, *2020 IEEE International conference for innovation in technology (INOCON)*, IEEE, pp. 1–4.

Faujdar, N., Saraswat, S. and Sharma, S. (2023). Human pose estimation using artificial intelligence with virtual gym tracker, *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, IEEE, pp. 1–5.

Gadhiya, R. and Kalani, N. (2021). Analysis of deep learning based pose estimation techniques for locating landmarks on human body parts, *2021 International Conference on Circuits, Controls and Communications (CCUBE)*, IEEE, pp. 1–4.

Kotak, D., Desai, J., Shah, R., Goel, P. and Patel, K. (2022). Yoga pose classification using angle heuristic approach, *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 1709–1714.

Liaqat, S., Dashtipour, K., Arshad, K., Assaleh, K. and Ramzan, N. (2021). A hybrid posture detection framework: Integrating machine learning and deep neural networks, *IEEE Sensors Journal* **21**(7): 9515–9522.

Mili, R., Das, N. R., Tandon, A., Mokhtar, S., Mukherjee, I. and Paul, G. (2022). Pose recognition in cricket using keypoints, *2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, IEEE, pp. 1–5.

Min, Z. (2022). Human body pose intelligent estimation based on blazepose, *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, IEEE, pp. 150–153.

Mitrović, K. and Milošević, D. (2022). Pose estimation and joint angle detection using mediapipe machine learning solution, *Serbian International Conference on Applied Artificial Intelligence*, Springer, pp. 109–120.

Mohammed, S. W., Garrapally, V., Manchala, S., Reddy, S. N. and Naligenti, S. K. (2022). Recognition of yoga asana from real-time videos using blaze-pose, *International Journal of Computing and Digital Systems* **12**(1): 1304–1295.

Mroz, S., Baddour, N., McGuirk, C., Juneau, P., Tu, A., Cheung, K. and Lemaire, E. (2021). Comparing the quality of human pose estimation with blazepose or openpose, *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*, IEEE, pp. 1–4.

Munea, T. L., Jembre, Y. Z., Weldegebriel, H. T., Chen, L., Huang, C. and Yang, C. (2020). The progress of human pose estimation: A survey and taxonomy of models applied in 2d human pose estimation, *IEEE Access* **8**: 133330–133348.

Nagarkoti, A., Teotia, R., Mahale, A. K. and Das, P. K. (2019). Realtime indoor workout analysis using machine learning & computer vision, *2019 41st Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, pp. 1440–1443.

Palanimeera, J. and Ponmozhi, K. (2021). Classification of yoga pose using machine learning techniques, *Materials Today: Proceedings* **37**: 2930–2933.

Rahman, Z. U., Ullah, S. I., Salam, A., Rahman, T., Khan, I., Niazi, B. et al. (2022). Automated detection of rehabilitation exercise by stroke patients using 3-layer cnn-lstm model, *Journal of Healthcare Engineering* **2022**.

Shah, D., Rautela, V., Sharma, C. et al. (2021). Yoga pose detection using posenet and k-nn, *2021 International Conference on Computing, Communication and Green Engineering (CCGE)*, IEEE, pp. 1–4.