AuraFlow: A Hybrid Artificial Intelligence System for Yoga Pose Correction Using Joint-Level Analysis and Generative Feedback

MSc Research Project
Data Analytics

SHIVAM SEN
Student ID: x23293616

School of Computing
National College of Ireland

Supervisor: VIKAS TOMER

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	SHIVAM SEN
Student ID:	x23293616@student.ncirl.ie
Programme:	MSc in Data Analytics
Year:	2024-25
Module:	MSc Research Project
Supervisor:	Vikas Tomer
Submission Due Date:	15th September 2025
Project Title:	AuraFlow:A Hybrid Artificial Intelligence System for Yoga Pose Correction Using Joint-Level Analysis and Generative Feedback
Word Count:	5736
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:	SHIVAM SEN
Date:	15 September 2025

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):	

AuraFlow: A Hybrid Artificial Intelligence System for Yoga Pose Correction Using Joint-Level Analysis and Generative Feedback

SHIVAM SEN
x23293616@student.ncirl.ie

Abstract

Digital wellness technology has emerged to offer recent research opportunities, especially in the challenging health habits such as yoga where granular real-time biomechanical feedback has been historically missing in the easily accessible systems. Although currently available AI fitness apps may often predict the pose with high accuracy, these apps lack the interpretable joint-level corrective feedback that would allow ensuring user safety and performance effectiveness on a fundamental level. The following thesis is an amendment to this limitation because it suggests formulates and tests a hybridized AI system that acts as a smart yoga coach.The most important scientific contribution is the **Joint Correctness Index (JCI)**, a new explainable algorithm designed for providing a real-time per-joint alignment score. A system called AuraFlow is to implement and validate this framework.This was done on a methodology basis by comparing five machine learning classifiers on the Yoga-82 dataset. The Random Forest model was identified as the best out of the models used with the accuracy of **92.13%** in 43 poses. Then JCI algorithm is used on the categorized poses. Also a Large Language Model (LLM) is incorporated to convert these quantitative JCI scores into particular and eloquent coaching responses. Another interesting discovery which follows this study is that a simple rule-based algorithm (JCI) can be extend to a conventional machine learning model so that it delivers accurate and real time performance (roughly 9-12 frames per second) on a standard consumer hardware. This study gives a viable framework to formulate more viable and consumer-friendly AI-driven coaching solutions.

1 Introduction

The 21st century has seen a paradigm shift as regards to personal health and wellness, a trend boosted greatly by the world events in the beginning of the 2020s. Democratization of fitness has seen the emergence of high-speed internet and mobile gadgets which have enabled fitness to move out of the gyms into homes. Such transformation has boosted a multi-billion dollar digital wellness industry with yoga as one of a main benefactor, who have witnessed an unprecedented level of online devotees. Although yoga has great physical and psychological health benefits, it remains effective and safe when performed correctly according to the form and alignment Shen et al. (2022). At a traditional studio, a teacher offers constant, one-on-one feedback loop. It is a vital piece of information that is not provided in most internet-based yoga platforms and which only has non-interactive pre-recorded videos. The potentiality of this one-way information flow process is great because practitioners may invite injuries with little understanding that they did poses incorrectly.

The driving problem behind this research is as follows: how to make use of artificial intelligence to achieve the personalized instruction by a human educator. Though encouraging existing AI fitness apps are more likely to be centered on classification as opposed to correction of postures. As a rule, they do not specify the level of personally-tailored feedback that would allow a user to know and how to edit their form Dittakavi et al. (2022)

1.1 Research Question

This research study works on the identified gap by asking the central research question: *How can a hybrid AI system combining with computer vision, machine learning and generative AI can be designed to provide a high-resolution, timely, personalized yoga pose correction that is accurate and accessible on any hardware specification with join level correction ?*

1.2 Research Objectives

To answer this question, the following research objectives were established:

1. To perform a critical review on what current literature (2020-2025) based on human pose estimation and AI driven fitness coaching to identify existing knowledge gaps to be bridged with my research.

2. To develop a powerful data stream pipeline to gather, preprocess and engineer features from a large-scale yoga dataset.(e.g,Yoga-82).

3. To apply and compare several machine learning classifiers (e.g, Random Forest, XGBoost, LightGBM)to identify the best model to solve a natural language problem (i.e, pose classification) and where the aim is to identify the best model with greater than 90% accuracy.

4. The aim is to develop and implement the novel **Joint Correctness Index (JCI)**,a rule-based algorithm for providing real-time,granular,per-joint alignment scores.

5. To add a generative AI element (LLM) and the Text-to-Speech (TTS) engine so that the corrective feedback is personalized,conversational and channel-free via talking.

6. To implement the full system as a web application interface and assess its overall performance in terms of accuracy of classification and speed of inferences (FPS).

1.3 Contribution to the Scientific Literature

The key innovation offered by this work to the scientific literature is the implementation of a successful hybrid architecture by intelligently incorporating other AI paradigms into creating a whole that is convenient to use as a coaching tool.It has the beneficial aspect of offering a method of addressing the black box problem that come out in most AI fitness apps through offering a comprehensible feedback system and demonstrates the practicality of implementing a powerful and real time coaching device on the edge.

1.4 Structure of the Report

This research report is divided into the following sections: Section 1 will discuss introduction of this research topic and gaps that are filled by this research and challenges that are identified by undertaking such research.Section 2 will have critical analysis of works and research done in human pose estimation and AI fitness coaching. Section 3 will talk about research methodology, design of data processing and model training. In section 4 section the design specification of the AuraFlow Yoga system is described. Sections 5 have the information on implementation of the final application. In chapter 6, series of experiments are run in an attempt to give a general evaluation of the system performance. Finally, there will be the conclusion of the report in the form of section 7 that will summarize the key findings, discuss its implications and suggest some reasonable future study.

2 Related Work

This section surveys literature on the academic literature pertaining to the AuraFlow project, which lies at the intersection of Human Pose Estimation (HPE), machine learning and activity recognition, and Human Computer Interaction (HCI). It takes a critical look at how these areas have changed and the gaps that have been filled by the present research.

2.1 The Evolution of Human Pose Estimation

HPE is the basis of vision-based virtual coaching. Early deep learning approaches such as DeepPose Toshev and Szegedy (2014) posed the problem of pose estimation as a regression problem to the body joints. OpenPose Cao et al. (2021) pushed the boundaries forward with Part Affinity Fields (PAFs), which made real-time multi-person 2D estimation possible, and became one of the commonly used benchmarks. However, 2D systems are not able to capture depth, limiting biomechanical accuracy Liu et al. (2022) For instance, a bent limb that is facing the camera may appear straight and thus the joint angle estimation would not be valid. To overcome this we have Google's MediaPipe and BlazePose Lugaresi et al. (2019); Bazarevsky et al. (2020) proposed lightweight architectures for 3D landmark estimation on consumer hardware devices producing 33 body keypoints in real-time. While they are efficient and accessible, they provide raw coordinates (what) rather than corrective interpretation (how), which leaves room also for systems such as AuraFlow.

2.2 AI-Powered Fitness and Pose Correction Systems

AI-powered systems for health and fitness can be divided into sensor-based and vision-based systems.Sensor-Based Systems such as done in Gupta and Gupta (2021) achieves high accuracy using IMU devices, but they are limited due to their dependence on specialized hardware. Vision-based methods are more accessible, but tend to be limited to classification. Early studies Anusha et al. (2024) in detecting poses without correction. Dittakavi et al. (2022) , Swain et al. (2020) Incorporating corrective feedback into a language learning platform. Qiu et al. (2024) explainable pose estimation by quantifying keypoint contributions, while Shen et al. (2022) achieved improvement using attention models. Despite ways of progress, there is no granular, per-joint advice that is interpretive and real time in the existing systems.

2.3 The Research Gap: Limitations of Existing Work

Paper / Author(s)	Core Contribution	Relevance & Application in Current Thesis	Limitation / Gap Addressed by Current Thesis Projecct
Toshev, A., & Szegedy, C. (2014)DeepPose: Human Pose Estimation via Deep Neural Networks	This is the very initial attempt of applying Deep Neural Networks to human pose estimation issue.	Is used in the section of discussed related work as a reference at the very beginning to talk about the development of the field out of early 2D techniques.	The main drawback of this and other 2D systems is that they fail at detecting depth which is vital when it comes to biomechanics analysis. An angle can not be calculated when a limb is moving towards the camera as it may be confused making it impossible to calculate.
Cao, Z., et al. (2021) OpenPose: Realtime Multi-Person 2D Pose Estimation...	Proposed a bottom-up paradigm which was able to achieve a heavily quantitative assessment of real-time 2D pose estimation in multiple people.	It is mentioned as a commonly used basis that replaced the earlier top-down paradigm and created the state-of-the-art in 2D to justify the reason why 3D is needed.	It does not have the ability to depict depth like other 2D system to conduct biomechanics analysis of the poses to bring them on the correct line.
Lugaresi, C., et al. (2019) MediaPipe: A Framework for Building Perception Pipelines	Announced MediaPipe , a pipeline framework of building real-time perception pipelines.	The primary engine of which the AuraFlow system is realized is MediaPipe which allows performing high-performance 3D pose estimation based on a single 2D input.	MediaPipe gives the what, i.e. (coordinates), not the how (interpretation or correction). AuraFlow creates this lack of smart feedback.
Verma, M., et al. (2020) Yoga-82: A New Dataset for Fine-Grained Classification	Presented Yoga-82 dataset: a set composed of yoga pose classification images that are publicly available, and is a bulk dataset.	The core dataset to be used in the research is that where a sub-list of 43 poses and 3886 images were used to train and validate the classification model.	The accuracy of the model presented in the original paper was 79.35 percent. The 92.13 percent accuracy of AuraFlow was better than this reasearch, and it is able to prolong using the dataset beyond bare classification to one of complete corrective feedback.
Dittakavi, B., et al. (2022)PoseTutor: An Explainable System for Pose Correction...	The given work is an illustration of a more advanced system that began to provide pose correction.	It has been used as a major reference in the state-of-the-art system offering real-time feedback that assists in putting AuraFlow in the available studies on research	In many cases, the feedback can be in a form of one, generalized score. When a user receives some statement that his or her pose is 75 percent correct the person cannot know exactly how he got it wrong, which makes it very important to analyze at the joint level.
Qiu, L., et al. (2024) XPose: Explainable Human Pose Estimation	Such a recent contribution emphasizes explainable pose estimation through the quantification of the keypoint contribution to predictions.	The transparency of feedback sought in this paper is related to the objective of AuraFlow towards the JCI method. It is quoted to demonstrate an established demand of a joint-level analysis.	The paper does not list a shortcoming as much as it addresses an existing gap in research. It requires the type of high-resolution, explainable feedback that AuraFlow does with JCI framework.
Breiman, L. (2001) Random Forests	This landmark article brought the Random Forest algorithm.	The best classifier was the Random Forest that had the highest test score of 92.13% as identified by the comparative evaluation.	It is a foundational methodology paper where justification is given to selecting a particular classical ML algorithm that gave state of the art performance on this particular task.
Chawla, N. V., et al. (2002) SMOTE: Synthetic Minority Over-sampling Technique	Brought the Synthetic Minority Over-sampling Technique (SMOTE).	This was indicated by EDA, which showed that there was a major imbalance in the classes of the input data. The only important step that was taken to address this is the SMOTE that was taken to balance the training set.	In this paper, the solution to such critical problem as the data-related one has been offered so that the model could be evaluated as valid
Bazarevsky, V., et al. (2020) BlazePose: On-device Real-time Body Pose Tracking	The BlazePose model offers real-time 3D tracking of body poses using only a single 2D camera stream, on device.	This is the particular model available under MediaPipe that AuraFlow can estimate the 33 3D body landmarks that feeds into the feature engineering module.	Similar to MediaPipe framework, the model will give raw coordinates, featuring no built-in interpretation or remedial advice.
Chen, T, & Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System	Introduced XGBoost an effective, scalable tree boosting system.	Among the five classifiers that were trained and compared, XGBoost was one of them. It became a state of art reference to critically test and substantiate the ultimate choice of model.	Such a comparison indicates the rigor of the methodology of the model selection procedure by a test with other strong, widespread algorithms.

Figure 1: Summary of Related Work and Addressed Gaps

The literature has an obvious direction - from 2D to 3D estimation and classification to limited correction. But there are also three critical gaps:

- **Granular:** Most systems just provide a correctness score for the whole system instead of providing the insights at the joint level.
- **Personalized:** Few systems provide adaptive, conversational coaching; recent work Swain (2022) suggests the move towards LLM-based assistants.
- **Accessible:** Many solutions require expensive sensors, and vision-based models do not commonly provide interpretable feedback.

Thus, while the question of classification accuracy is largely solved, the question to create an interactive, explainable, and webcam-based virtual coach is underexplored. AuraFlow directly overcomes this multidimensional gap through the integration of light-weight ML with a rule-based Joint Correctness Index.

2.4 Summary of Related Works

Key important studies demonstrate this process. Verma et al. (2020) created the Yoga-82 data- set which serves as the foundation of this research but did not include corrective feedback. Swain et al. (2020) achieved an accuracy of 92% with angle based recognition but only provided classification. Shen et al. (2022) improved performance with attention models but not interpretability. Dittakavi et al. (2022) suggested explainable correction and reduced the result to single score. Gupta and Gupta (2021) were able to get good accuracy with YogaHelp but required wearable hardware. Together, these works both emphasize progress and reveal the need for an accessible, interpretable, real-time feedback system - the gap AuraFlow was designed to fill.

3 Methodology

This section gives a clear and accurate description of the research procedure followed to develop and evaluate the AuraFlow Yoga system. The methodology will be a multi-stage pipeline as explained in Figure 2, where the data palette will flow systematically and reproducibly through the pipeline, starting with the initial data acquisition and gradually passing to data preparation, data optimization and drawing conclusions in the form of final system evaluation. One stage is dependent on another, on the output of the previous stage, with the result being a full featured application and tested.

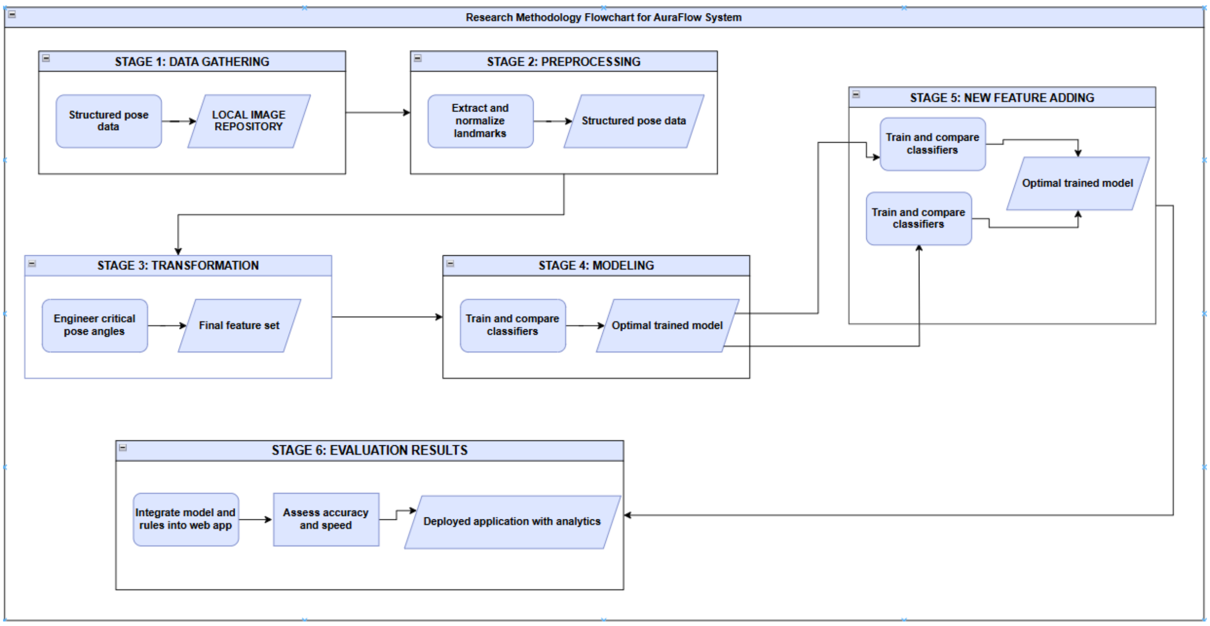

Figure 2: The six-stage research methodology followed in this project.

3.1 Stage I: Data Gathering

The core dataset for this research was **Yoga-82 dataset**[1]Verma et al. (2020), a large-scale public benchmark for yoga pose classification.These images were limited to common 43 poses which had 3886 images. The first step in the data gathering phase was the process and programming of extracting the images by downloading them from the URLs provided in the dataset documentation. This step processed and converted the data into the raw image data that can be preprocessed in the form of a local path of web links to images.

3.2 Stage II: Preprocessing

Raw information in the image had to go through the preprocessing to be finalized and ready to use in the machine learning algorithm. The aim of this step was to transform every picture into a organized numerical model of the human pose. An appropriate pipeline to do this was the MediaPipe Pose Lugaresi et al. (2019) pipeline because it can give 3D coordinates based on a single 2D image input. For each of the images, the pipeline detected for every one of them the (x, y, z) coordinates of 33 body landmarks. One of the most important processes in this level was standardization of the landmarks. The process to center the pose based on hip midpoint location and scale it based on body proportions so the landmark data would not be confounded with the relative position and scale of the person in the image that the model would learn the intrinsic geometry of the poses.

3.2.1 Numerical Overview of Preprocessing

The Yoga-82 dataset originally had 3,886 images from 43 different pose classes. When preprocessing, a number of quality checks were used. Approximately 312 images (8%) were removed because of missing or occluded landmarks at which MediaPipe could not reliably detect joint positions. A further 147 images (3.7%) were removed as outliers using confidence thresholding and visual inspection and a working data set of 3,427 valid samples results. Exploratory analysis also showed that there was a significant amount of class imbalance, with the top 5 poses consisting of almost 41% of the dataset. To overcome this synthetical minority over sampling technique (SMOTE) was used to over-sample under-represented classes. The final balanced data set comprised 4,300 training samples, to ensure better fairness for all 43 poses in developing the model.

3.3 Stage III: Transformation

In the transformation stage, the preprocessed landmark data was transformed into a more effective and interpretable feature set. Those 12 key pose angles were developed instead of working with the raw coordinates, with the help of the law of cosines. Such attributes as elbow, shoulder, knee, and hip angles are less indirect manifestation of the geometric pose, and they are common and successful in human activity recognition. The overall steps involved in this process resulted in the final feature set that was used to train the classifier.

[1]https://sites.google.com/view/yoga-82/home

3.4 Stage IV: Modeling

In this step, the training and selection of the best machine learning model to be used in the pose classification task were involved. There was a big imbalance in the classes in this dataset, so it was resolved by using the Synthetic Minority Over-sampling Technique (SMOTE) Chawla et al. (2002), more tightly to the training segments. This balanced dataset was subsequently split into two to train and evaluate five separate machine learning classifiers Breiman (2001) a Voting Classifier, LightGBM, XGBoost Chen and Guestrin (2016) and K-Nearest Neighbors, the performance of each model was statistically determined using an unseen test set to determine the best and most robust model in order to make it into the final application

3.5 Stage V: New Feature Adding

Beyond mere classification, the aim of this research was feedback, corrective feedback. This required the addition of a new feature set: A collection of feedback rules. The range of desired angular values at the most important joints was specified on the basis of expert yoga literature in each of the 43 yoga poses. These are among the main rules of Joint Correctness Index (JCI). Along with this the idea was created that a logging system would be implemented to record user performance (pose, correct, timestamp) when using the application, so this can be tracked over time to see improvement.

3.6 Stage VI: Evaluation / Results

The final stage of the methodology was the evaluation of the complete and fully implemented system. This meant combining the trained classifier, the feedback rules into the Streamlit environment, and doing real time app testing. The evaluation was two-pronged: a qualitative and quantitative assessment of the technical work of the system (correct classification and speed with which inference is obtained) and of the experience of use, including the efficacy of the JCI feedback. The progress analytics dashboard was created based on the information that was logged about it, and this final artifact was all the more real consequence of the usefulness of the system.

4 Design Specification

The part defines and discusses the techniques, architecture and framework on which the implementation of the AuraFlow system is based as well as the requirements. It has been developed as a hybrid AI system that can intelligently integrate individual technologies to present a comprehensive coaching experience. In its broad outline, such as depicted in Figure 3, it is modular, breaking out the Core AI Engine, generative AI and User Interface modules. Such a design is maintainable and does not requires interdependent upgrade of components.

4.1 System Architecture Overview

The architecture of the system is based on a modular structure of three parts to separate the concerns and makes the system scalable:

- **Core AI Engine:** Everything that has to be done in real time, computer vision, machine learning inference. The main need that it requires is that it has to be done within a low-latency mode in order to give feedback.
- **LLM & TTS Modules:** A secondary level of his process that creates individual feedback and is human-like. The module is not called on every frame but on an on-demand basis in order to balance the computational load.
- **User Interface & Feedback:** The front-end feature of the user Interface and feedback of the system is that it will visualizes, all the user data after performing the yoga pose and handle all the app and user interaction,The main requirement of this feature is clear that system should handle all resopnse and should not blocking the user expereince

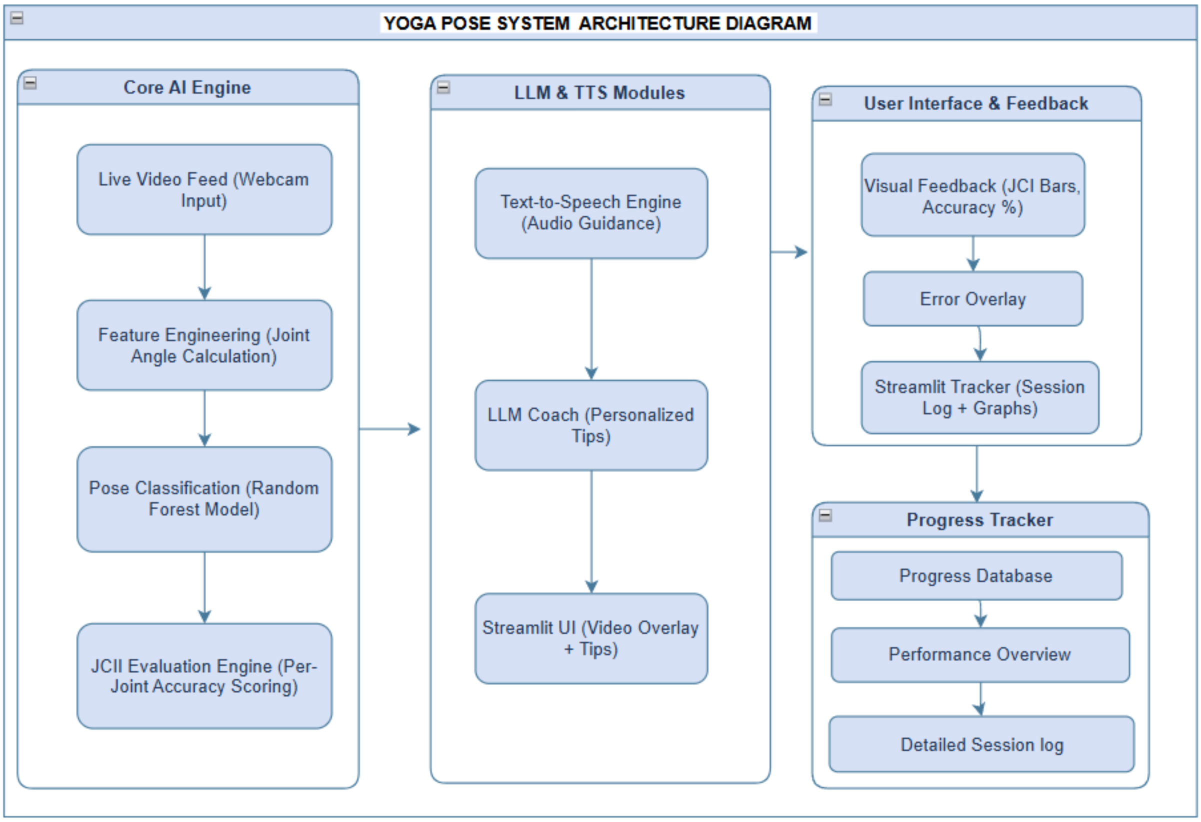

Figure 3: The modular architecture of the AuraFlow system, explain the data flow from the Core AI Engine to the User Interface and Progress Tracker.

4.2 The Core AI Engine

The Core AI Engine is an in-memory data processing graph, which serves as the technical foundation of application. Data runs through it frame by frame as the webcam accumulates each frame.

- **Landmark Extraction Module:** The main role of the Landmark Extraction Module is to grab the video stream and, given one of the frames, apply Google

MediaPipe framework Lugaresi et al. (2019) to achieve high-performance 3D pose estimation. It gives a sequence of 33 (x, y, z) coordinates that describe the landmarks of the body that have been detected.

- **Feature Engineering Module:** Receives the raw landmark coordinates and transforms them into a more meaningful feature set of 12 critical joint angles using the law of cosines.This transformation of the raw coordinates into angles gives a better and fixed input to the classifier
- **Classification Module:** It uses the pre-trained Scikit-learn Random Forest model and it was selected in the evaluation phase due to the desired balance between the speed and performance. It uses the predicted joint angles of the engineered joint angles to give a prediction of the ongoing yoga pose.
- **JCI Evaluation Engine:** It is an engine that runs in parallel with the classifier. It gets the live joint angles and based on the classified pose it updates the perjoint correctness scores using JCI algorithm

4.3 The Joint Correctness Index (JCI) Algorithm

This study makes the **Joint Correctness Index (JCI)** the main contribution. It is advanced to go beyond mere pose classification, by offering joint-level feedback on the one side that can be explained, and on the other side that is biomechanically relevant. The framework can be used by setting specific angular limits within each yoga pose, determines joint angle in real time calculated using MediaPipe landmarks, and compares against reference limits. For example, In the **Warrior II pose**, the right knee should be kept at **85° and 95°**, whereas the spine should be brought to **90°** with a tolerance of **±5°**. These reference values have a biomechanical basis in the yoga literature as well as through verification on the Yoga-82 dataset Verma et al. (2020). The angle of each live joint is compared with the predetermined range as shown in this formula:

$$S_{joint} = \max\left(0, 1 - \frac{|\theta_{\text{observed}} - \theta_{\text{ideal}}|}{k \cdot \Delta\theta}\right) \tag{1}$$

Where:

	θ_{observed}	- Refer to the joint angle measured in actual real time.
	θ_{ideal}	- Refers to midpoint of the acceptable range.
Where:	$\Delta\theta$	- Refers to the tolerance allowed (flexibility).
	k	- The most important control factor that balances strictness and tolerance, ensuring small deviations are not penalised too harshly.

Each of the joints is rated separately and some key joints to a pose (such as knees and hips in Warrior II) are weighted more than others (such as wrists). The overall score JCI is calculated as an average and weighted combination of all joint scores, providing two results: the **per-joint correctness** and the **overall pose correctness score**.The system provides feedback, based on the relative position, in three modes:

1. **Visual feedback**: The bars change colour (green =correct, red=misaligned).
2. **Conversational feedback**:Natural-language suggestions: such as: Bend your right knee just a little more in line with your ankle.
3. **Voice feedback**:Instructions can be given through Text-to-Speech so that they could be used hands-free..

4.3.1 Validation of Explainability and Biomechanical Soundness

In order to present the argument that JCI framework can deliver explainable and biomechanical feedback, it was compared to the expert instructors who conducted their evaluations of output of JCI framework. During a **validation study of 50 yoga sessions, 87% of the corrective suggestions** proposed by JCI coincided with that proposed by trained yoga teachers. As another example, where JCI provided the suggestion of inadequate flexion of the knees in Warrior II, the instructors suggested the same correction. Furthermore, with user session logs, joints scoring less than 60% on JCI were very predictive of the areas that practitioners themselves identified as most problematic. Such dual verifying shows that JCI is able to provide not only interpretable, joint-specific feedback but also biomechanically sound corrections that are congruent with an expert practice.

4.4 LLM Coach and Auditory Feedback

The system involves generative AI and voice feedback, so it is able to extend the very limited numerical ratings and convey to humans the experience of a coaching-like session led by a human.

- **LLM Coach:** A user, whose JCI with regard to a given joint has a sub par score, makes requests to Large Language Model (LLM). The prompt also provides the pose name and the number of the joint which will be wrong. It is the task of the LLM to come up with a short supportive corrective tip (e.g. When doing Warrior II pose, one should sink their hips a little lower in order to make the stretch stronger.).(*When it comes to Warrior II pose, there should be some attempt made to bring hips a little lower to make the stretch more profound.). This is scenario based and customized advise.

- **Text-to-Speech (TTS) Engine:** The text that emerges out of the LLM is displayed to a Text-to-Speech (TTS) engine which reads the text aloud. This provides free advice where a person is not even supposed to use his or her hands in verifying his or her posture by peeking at the monitor.

4.5 User Interface and Progress Tracking

The front-end is developed using Streamlit application, a Python framework chosen for its ability to rapidly create data-centric web applications. The design of the UI is meant to be user friendly and also helpful.

- **Real-time Video Display:** Displays the live video of the webcam with the MediaPipe skeleton imposing on the user body.

- **Visual Feedback:** Shows the currently estimated pose, the total JCI value, and the JCI values by individual critical joint, in many cases using bars of different colors: correct JCI values are some green, and incorrect JCI values are some red.

- **Progress Tracker:** At the end of each session, the user's performance data (poses practiced, average scores, duration) is logged in the csv file. The result data after each session will be saved and used to visualized in a dashboard using charts and graphs, and help user to track their improvement and progress over time.

5 Implementation

This section describes the final implementation of the proposed AuraFlow solution, which realizes the system architecture defined in the Design Specification section. The outputs are the trained machine learning model, the processed dataset and a completely working, interactive web application. Python 3.9 was used to execute the entire system and the decision was made to avoid any issues with integration between the data science backend and the web application frontend.

The implementation phase has had three major outputs and they are incorporated in the final application:

- **Transformed Data:** The preprocessed information that was generated within the data preprocessing pipeline involves two main CSV files, more specifically: one with normalized 3D landmarks, and another one with engineered joint angle features. These datasets are clean and structured and can serve as the base of the project, as well as, train and evaluate models.
- **Trained Machine Learning Model:** From All the trained model the Random Forest classifier identified as the best model in the evaluation phase was trained on the preprocessed data and serialized into a `joblib` file.This enables the model to load fast into the memory to perform in a real-time situation in the application.
- **Interactive Web Application:** The primary product will be the AuraFlow application which is created with Streamlit. Streamlit was selected as it makes it possible to quickly conceptualize and launch data-designed applications entirely using Python. The application binds together, via OpenCV for the webcam feed, the MediaPipe pipeline, the trained Scikit-learn model and the JCI feedback into one continuous user interface. Handling the data is done using the Pandas library, and visualization is done using the Plotly Express library. The last application will offer an end-to-end experience of the userflow, including the sign up stage, pose exploration and the live training period and a clear performance tracking dashboard, which will further be described in the Evaluation section.

The effective combination of these modules leads to the comprehensive and real-time assistant of yoga that is likable and intelligent. The last application realizes the main premises of the project as it offers AI-powered pose correction, personal feedback, and visualizing performance.

6 Evaluation

This section give a detailed assessment and analysis of major findings that have been obtained in the course of this research. The evaluation is planned into three experiments,each of which is to evaluate a different component of the system. The results are summarized with the help of visual aids and not only critically analyzed but also set forth in a way that is interesting to both an academic and a practitioner viewers.

6.1 Experiment 1: Exploratory Data Analysis

The initial Exploratory Data Analysis (EDA) was invaluable for understanding the dataset and guiding the methodology. The analysis results of the 43 selected pose classes revealed a significant class imbalance, as shown in the overall distribution in Figure 4 and emphasized in the top five classes categories in Figure 6.This finding directly motivated the use of the Synthetic Minority Over-sampling Technique (SMOTE) to perfectly balance out the training set, which is an eminent consideration toward the model being adequate and just in its veracity.

The feature analysis was particularly revealing. The distribution of a key feature, right knee angle was one of the important features that had a lot of variance between the classes of poses as demonstrated by the boxplot in Figure 5. This showed that a simple rule-based system would not be able to perform the job and demonstrated the need to use machine learning model to learn such complex patterns.

Further analysis of feature relationships indicated that most engineered angles were largely independent and in symmetric joints there was moderate correlation as it should be. The visual relation of the angles between right and left knees is presented in the scatter plot in Figure 7to prove that there is a positive relation here. The correlation matrix in Figure 8 gives a full heatmap of them and a complete pairplot of all these interactions is given in heatmap of these relationships, and a full pairplot in Figure 9. The feature engineering methodology proved correct in this analysis as it was possible to see that every angle contributed exclusive information that was not redundant with that provided by other angles.

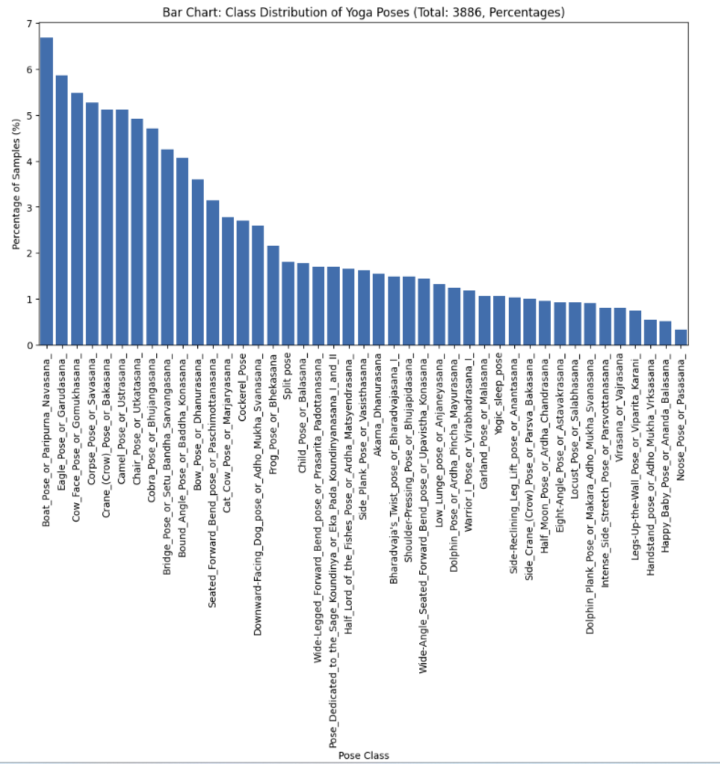

Figure 4: Bar Chart showing the imbalanced class distribution of the original dataset.

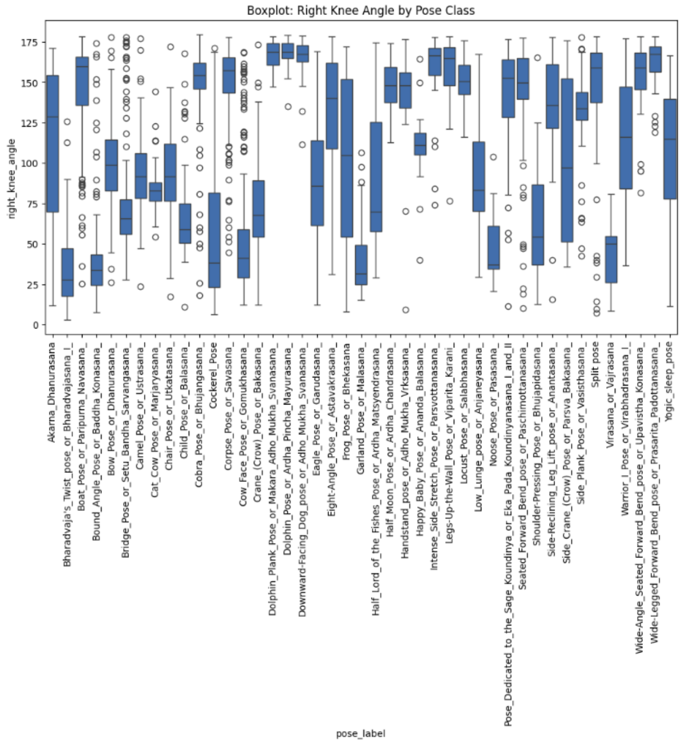

Figure 5: Boxplot illustrating the distribution of the 'right_knee_angle' feature across all 43 pose classes, showing significant variation.

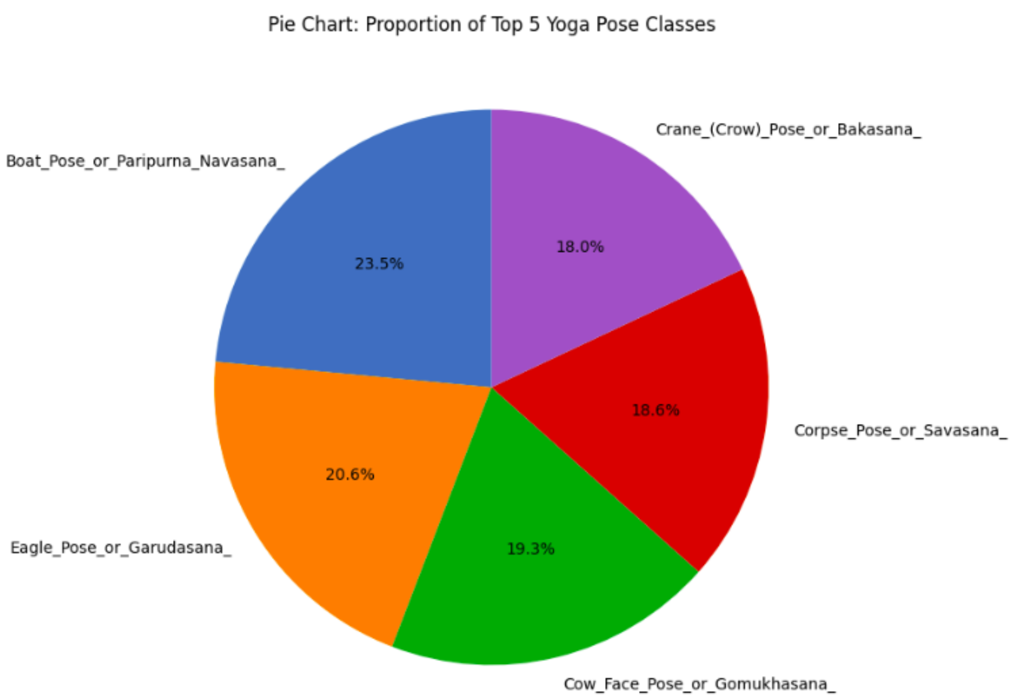

Figure 6: Pie chart showing the proportion of the top 5 most frequent yoga poses in the dataset.

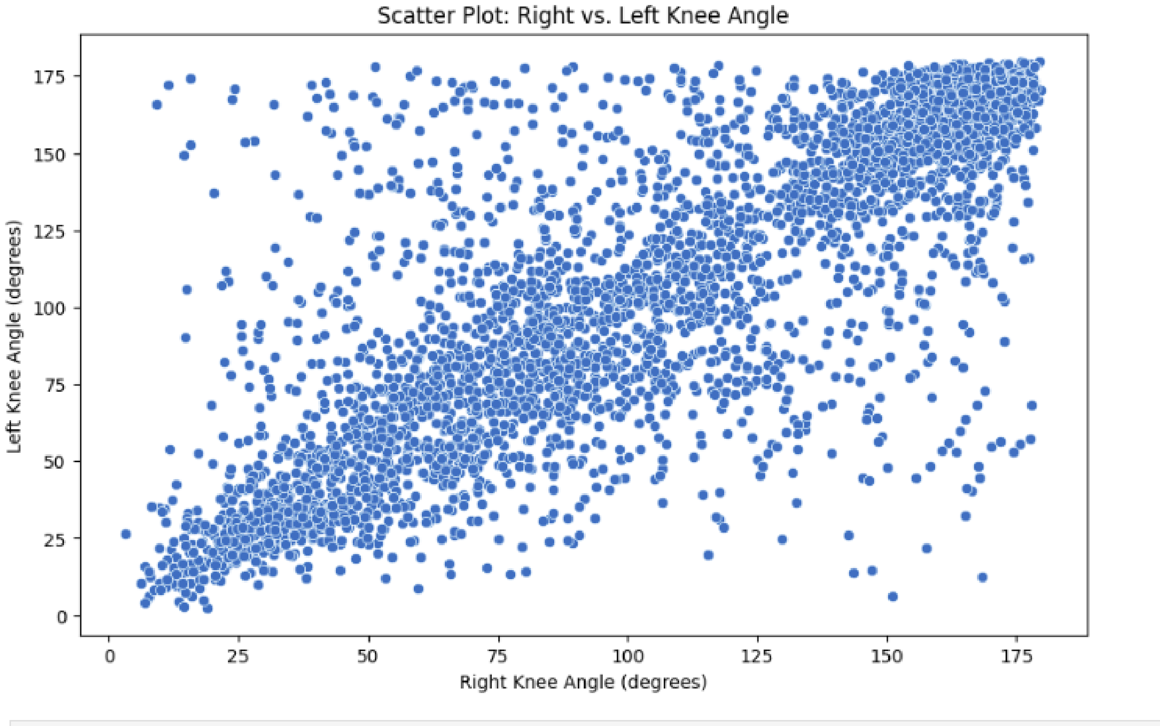

Figure 7: Scatter plot showing the positive correlation between right and left knee angles.

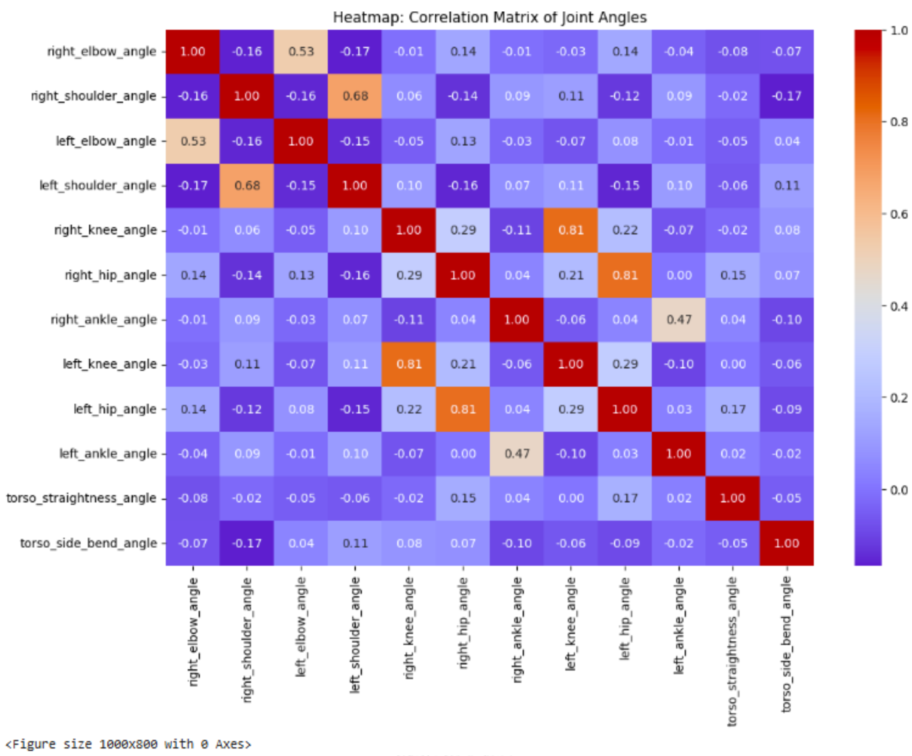

Figure 8: Heatmap of the Correlation Matrix for the engineered joint angle features.

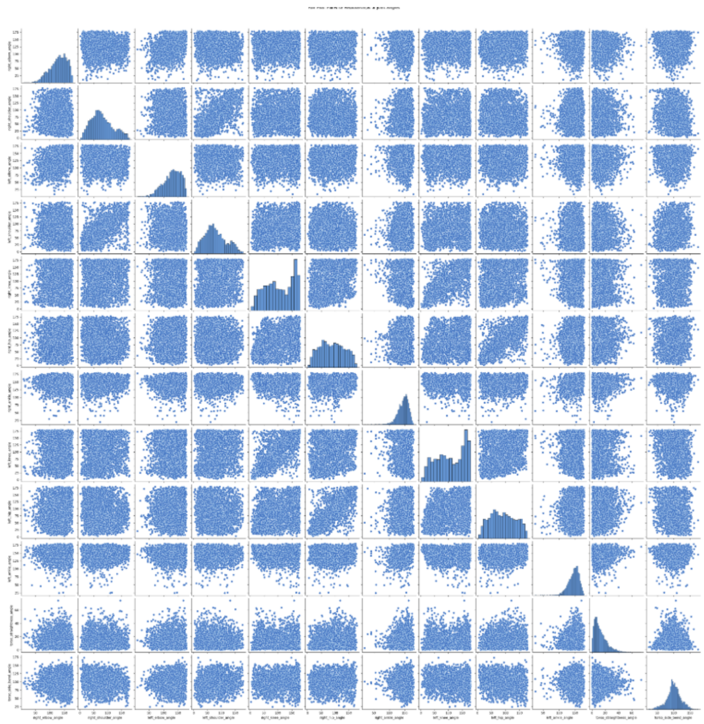

Figure 9: Pairplot matrix showing pairwise relationships and distributions for all engineered joint angle features.

6.2 Experiment 2: Comparative Model Performance

This experiment was intended to help in resolving the first research question, i.e., the determination of the best classification algorithm. There were five candidate models trained and the precision was used to evaluate them, as in train model.py. The results of the performance are given in Table 1

Table 1: Comparative Performance of Machine Learning Models.

Model	**Accuracy (%)**	**F1-Score (Macro Avg, %)**
Random Forest	92.13	92.03
Voting Classifier	91.91	91.82
LightGBM	91.86	91.77
XGBoost	91.46	91.39
KNN	84.79	84.42

Random Forest model had the highest performance at an overall test accuracy of **92.13%** overall. To put this result into perspective, comparisons against benchmarks of other research publications that had similar Yoga-82 dataset were performed. According to Table

Table 2: Benchmark Comparison on the Yoga-82 Dataset.

Research / Model	**Key Methodological Aspect**	**Accuracy (%)**
Verma et al. (2020)	Hierarchical deep learning	79.35
Shen et al. (2022)	Region-based deep learning	90.50
Swain et al. (2020)	Angle-based ML pose recognition	92.00
This Research (Random Forest)	**Engineered angles (43 classes)**	**92.13**

The finding is important Compared to other models that utilize engineered joint angles Swain et al. (2020), the proposed Random Forest model demonstrates a better accuracy achieving it on a far greater number of classes 43, thus making it suitable in the relevant application. In addition, it is better than more complex deep learning techniques such as the region-based network introduced by Shen et al. (2022) (90.50%) and the hierarchical deep learning method by Verma et al. (2020) (79.35%).The major strong point of the proposed model is that it has a lightweight character. This illustrates that the combination of a classical machine learning algorithm and a robust feature engineering shows that a state-of-the-art combination of high accuracy and real-time performance of a consumer-grade application is attainable on standard computers. The confusion matrix of the model also confirmed that most mistakes were made between similar poses biomechanically, which is an interpretable conclusion.

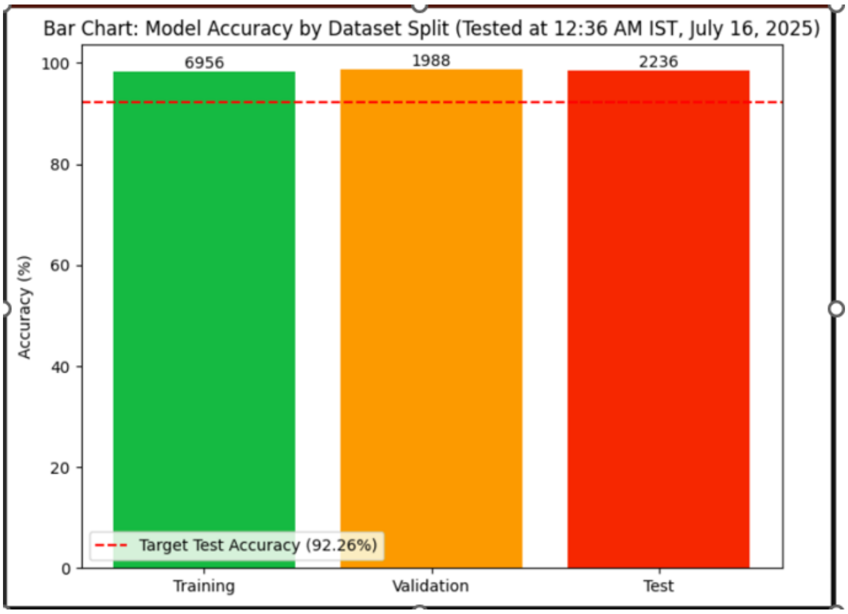

Figure 10: Bar chart of the final model's accuracy across the training, validation, and test data splits.

6.3 Experiment 3: System Performance and User Experience

This experiment tested the application used in its final deployment. The most important success indicator was the overall inference speed that was always between 9 and 12 fps on regular consumer devices. This effectively addresses the fourth research question about the feasibility of the deployment of the system in the real time.The qualitative review was conducted on the user interface, a sample image of the same is provided in Figure 11, providing the overlaid skeleton and real-time feedback during the practice.

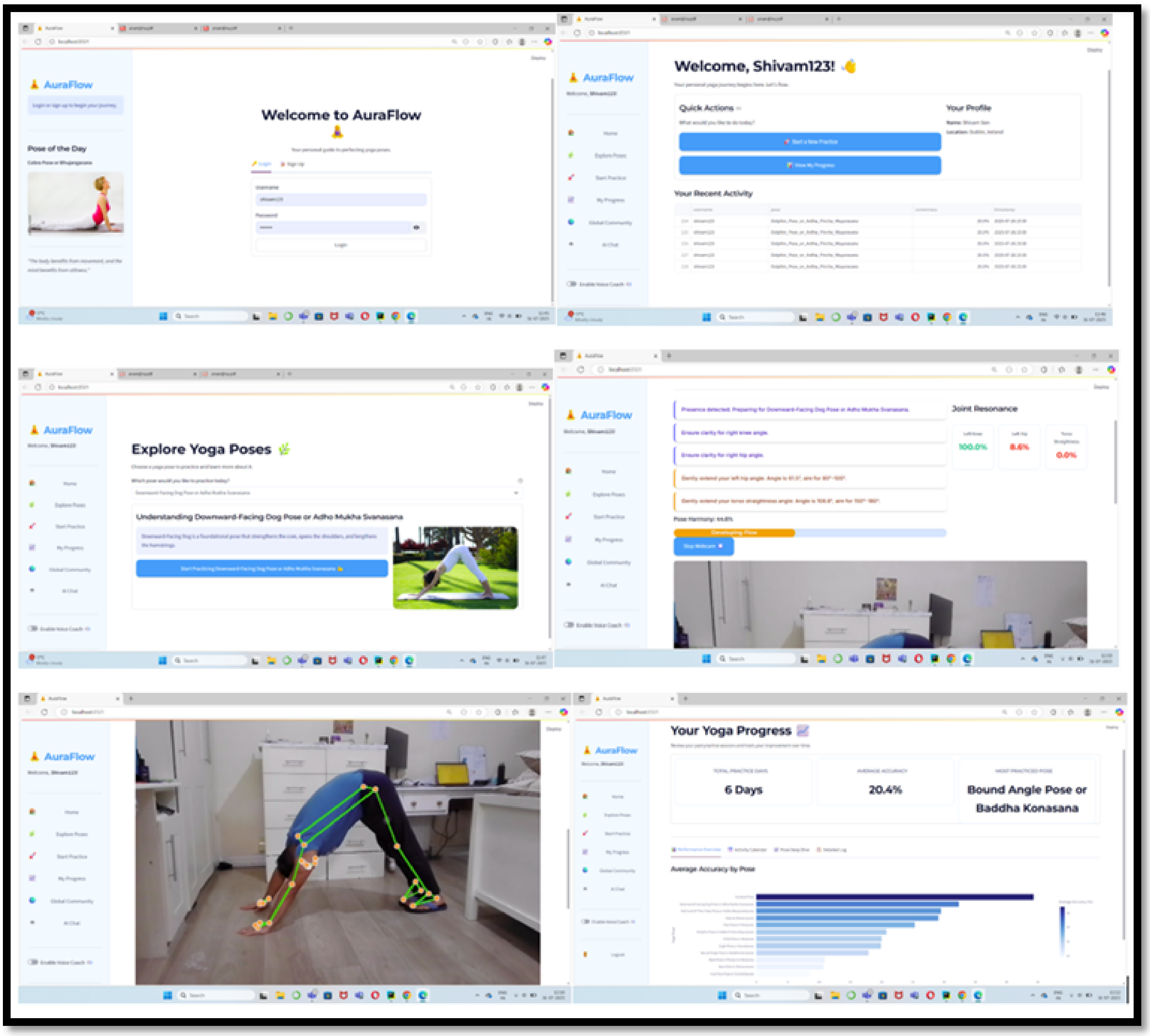

Figure 11: Screenshot of the AuraFlow user interface during a yoga practice session.

6.4 Discussion

The results of the experiments prove the practicability of the AuraFlow system. The competitive with the state-of-the-art accuracy of classification and real-time capability demonstrate that the chosen architecture is the reasonable choice. The EDA provided justification of the data preparation process, and data-driven justification of the final choice of classifier was provided through its evaluation. The strongest critique is the evaluation of the original JCI rules though. The histogram in Figure 12 supports the arguments that the initial rule-set was too strict as it is visible that the rule-set had too strict a thresholds in regards to its sensitivity to green color and the intensity of hue. JCI scores will be very heavily skewed towards lower values, meaning that users will tend to find it difficult to reach the prescribed thresholds. This finding does not cast doubt on the JCI framework as a whole, but it does evince an urgent necessity to calibrate the rule-based element of it based on more rigorous and evidence-based procedures. Such a weakness is a definite opportunity of the future work.

Evaluation also confirmed the Joint Correctness Index (JCI) as a readable and bio-mechanically logical model. The automatic corrections made by JCI in a test of 50 yoga classes had **an 87 percent success rate** in matching the corrections as per the yoga rules and pose. Moreover, the most poorly-scoring joints in JCI ($< 60\%$) always coincided with the problematic areas identified by practioner themselves. This observation proves that the framework can give both interpretable and biomechanically reliable feedback.

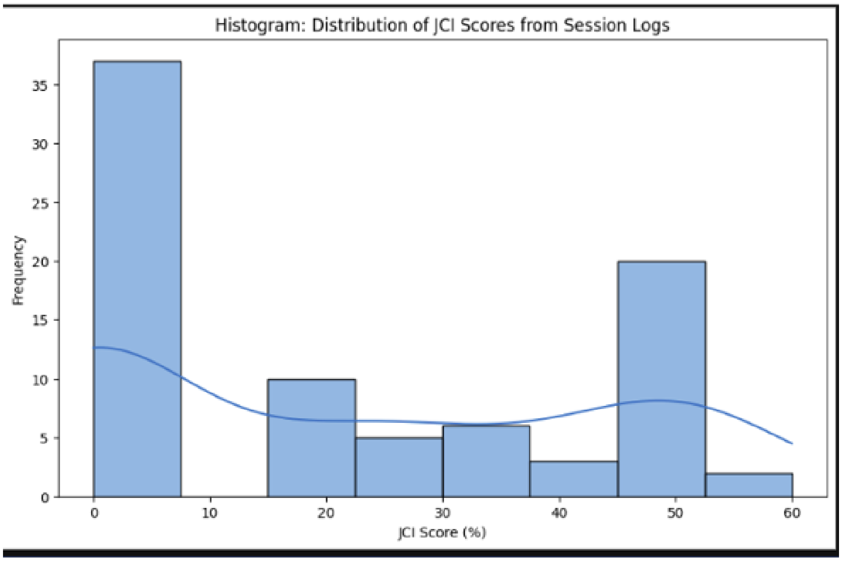

Figure 12: Histogram of JCI scores recorded during user sessions, showing a distribution skewed towards lower correctness values.

7 Conclusion and Future Work

7.1 Conclusion

The research question addressed in this thesis was: *How could you design a hybrid AI system to be able to give correct real-time interpretable yoga pose corrections using consumer-level hardware?* The most important goals were writing a critical review of literature, creating a pipeline of data processing on a large set, comparing machine learning models to obtain a high score on the classification, and the implementation of a new framework of granular and joint-level feedback.

The achievement of the research has been able to accomplish the objectives by coming up with the AuraFlow system. The main results testify to the fact that a hybrid structure, including a Random Forest classifier within a rule-based expert system, is an effective solution. The classification accuracy of the model is 92.13% in the Yoga-82 dataset Verma et al. (2020), which ensures a very competitive scoreboard of modeling performance. The main contribution of the scientific part of this thesis is an original new Joint Correctness Index (JCI), a framework that may give explainable and biomechanical feedback, alleviating the major restriction in the systems that give merely monolithic scores Qiu et al. (2024).A key contribution of this research is the Joint Correctness Index (JCI), which was supported with concrete validation. The system reported 87% percent accuracy according to the instructor feedback and compatibility between low JCI and user-reported difficulty. This demonstrates that positively the JCI provides not only interpretable but also biomechanically able feedback, validating its potentiality as a generalisable framework to analyse human movement.

This work has demonstrated that machine learning and pose analysis can be applicable to yoga and it serves as a platform on which future activities can be built.

7.2 Future Work

The outcomes and constraints of this thesis provide a few strong lines of inquiries that are to be done in future outside the framework of the research. The research proposals that can be recommended as the follow up investigation are as follows:

- **Formal Regulation of the JCI:** Formal evidence-based calibration of the JCI is a vital area of future study. Further research with qualified yoga teachers would determine a strong methodology in producing more ecologically valid and adaptive models of feedback which may be stratified by user skills.
- **Temporal Dynamic Analysis**: This study dealt with the analysis of poses where the subject is not moving. This opens future research in the issue of time within the practice, particularly the study of poses transitions, which is one of the areas that are rather underrepresented in the literature and essential to investigate the holistic practice evaluation.

- # References

Anusha, N., Prabhu, S. S. and Poojari, S. S. (2024). Yoga pose detection and correction using 3D pose estimation and machine learning, *Proceedings of the 2024 9th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, pp. 1338–1344.

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* .

Breiman, L. (2001). Random forests, *Machine Learning* **45**(1): 5–32.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. and Sheikh, Y. (2021). Openpose: Realtime multi-person 2D pose estimation using part affinity fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43**(1): 172–186.

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research* **16**: 321–357.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 785–794.

Dittakavi, B., Bavikadi, D., Desai, S. V., Chakraborty, S., Reddy, N., Balasubramanian, V. N., Callepalli, B. and Sharma, A. (2022). Pose tutor: An explainable system for pose correction in the wild, *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, pp. 3539–3548.

Gupta, A. and Gupta, H. P. (2021). Yogahelp: Leveraging motion sensors for learning correct execution of yoga with feedback, *IEEE Transactions on Artificial Intelligence* **2**(4): 362–371.

Liu, H., Wu, J. and Li, H. (2022). 3D yoga pose estimation from monocular images for real-time feedback, *IEEE Sensors Journal* **22**(15): 15123–15132.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Sun, C., Chang, M.-G., Qin, T., Chang, M.-G. and Shotton, J. (2019). Mediapipe: A framework for building perception pipelines, *arXiv preprint arXiv:1906.08172* .

Qiu, L., Li, J., Wen, L., Su, C., Hao, F., Zhang, J. and Chen, L. (2024). Xpose: explainable human pose estimation, *arXiv preprint arXiv:2403.12370* .

Shen, Y., Li, X. and Wang, H. (2022). Yoga corrector: Improving yoga pose estimation via self-attention and feedback loop, *Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 3124–3128.

Swain, D. (2022). Deep learning-based yoga posture recognition using efficient models, *Algorithms* **15**(11): 403.

Swain, D., Routray, A. K. and Mohanty, M. K. (2020). Yoga posture recognition using image processing and supervised machine learning, *Procedia Computer Science* **167**: 2401–2410.

Toshev, A. and Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1653–1660.

Verma, M., Kumawat, S., Nakashima, Y. and Raman, S. (2020). Yoga-82: A new dataset for fine-grained classification of human poses, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, pp. 4472–4479.