# AI Powered Fitness Coach System Using Machine Learning for Personalized Workout

[1]Dr. Kavitha Subramaniam, : [2] K Ashvitha ,[3] E Gopika ,[4] R S Abinaya , [5]Mrs K Cinthuja

[1]*Professor of Department of Computer Science,and Engineering collage,Panimalar Engineering college* Chennai, Tamil Nadu

[2,3,4,5]Students of *Department of Computer Science and Engineering, Panimalar Engineering college* Chennai, Tamil Nadu

Email:

[1]kavitha.pec2022@gmail.com,[2]ashvitha202005@gmail.com,[3]egopika645@gmail.com,[4]rsabinaya02@gmail.com

***ABSTRACT : With the increasing power of artificial intelligence (AI) and machine learning (ML), the health and wellness sector has seen a major shift. Customized fitness solutions are now getting more real, addressing individual users' specific needs and objectives. This paper introduces a web-based AI-driven fitness coach system to develop personalized workout timetables through sophisticated machine learning algorithms. Central to the system is a LightGBM classifier that suggests customized exercise routines based on user inputs including type of workout, time, weight, and set order preference.***

***Through the power of machine learning, the system is able to automate and customize fitness planning, becoming efficient and scalable for various fitness levels. The system proved to be highly accurate in creating customized workout plans, emphasizing its capability to offer valid and easy-to-use health advice. This research enriches the growing body of AI-based fitness applications and provides a working and versatile solution to aid individual health management.***

***Keywords: Customized Workout Routines, LightGBM,***
***Feature Encoding, Numerical Data Processing,***
***Hyperparameter Tuning, Machine Learning in Fitness***

## INTRODUCTION:

With more individuals attempting to achieve their own personal fitness and health objectives, personalized fitness programs have become increasingly popular. Although offthe-shelf, generic training schedules have value, they are less likely to meet each individual's particular needs, whether these needs are along the lines of time, strength, or existing fitness level. This lack of personalization can create a lack of motivation, increased learning time, and continuation with routines that will not serve them well.

Artificial intelligence (AI) and machine learning (ML) technology offer a novel and innovative approach to providing extremely tailored exercise recommendations. AI already has a formidable influence in various industries, and the health and fitness sector is not an exception. AI machines can analyze gigantic sets of information, recognize trends, and suggest something based on feedback from the user — with some level of precision and responsiveness that basic systems can't aspire to match.

The next article suggests a web-based system that has the ability to automatically create personalized workout timetables through the use of machine learning algorithms. LightGBM, a high-performance machine learning algorithm, forms its core where it can efficiently deal with both categorical and numerical data. Due to this reason, it is highly appropriate for fitness applications where variables like workout type, duration, and weight need to be handled with accuracy. LightGBM's efficiency and high accuracy of predictions make it a great option for building a flexible, user-friendly system capable of offering reliable, personalized fitness recommendations.

The rest of this paper is structured as follows: We begin by summarizing prior work related to the use of AI in fitness systems. We then explain the methodology, including data preprocessing, model training, and evaluation. We then present the results, including model performance. Lastly, we summarize with key takeaways and present recommendations for future enhancements and expansions to the system.

## II. RELATED WORKS:

Yadav et al. [1] proposed a fitness application based on machine learning algorithms that would design customized diet and workout schedules according to individual users' dietary lifestyles and goals. Their solution targets making the predictions both reliable and accurate to assist users in not just planning workouts but also monitoring their progress. This study brings out how important technology is in contemporary fitness careers, particularly in providing personalized advice that assists the individual in achieving their health aspirations.

Likewise, Felix et al. [2] investigated the potential of Long Short-Term Memory (LSTM) models in enriching the gym experience by predicting the next word of workout descriptions in an Android app. The aim was to present real-time updates and dynamic workout descriptions aligned with users' fitness objectives. Remarkably, the model was able to attain 95% accuracy, demonstrating the potential of natural language processing (NLP) in contributing positively to user engagement and maintaining pace with workout routines.

Rehman and Kumar [3] approached the problem differently by integrating computer vision with fitness tracking. Their study entailed applying YOLO (You Only Look Once) and MediaPipe to identify workout poses. Through the detection of significant body points, they used ensemble learning methods to classify various exercises through images. Their findings indicated notable improvements in performance, exemplifying the application of sophisticated algorithms to create intelligent fitness apps that can accurately evaluate users' form and physical activities.

Following the same line, Rahman et al. [4] presented a smart system intended to monitor physical exercises automatically through machine learning methods. Their approach analyzes real-time video streams to identify and classify various exercises by detecting keypoints from body movements of humans. What is particularly helpful with this work is that it exceeds mere exercise logging — it also offers features such as automatic rep tracking and instant feedback, enabling users to maintain their workouts in line without constantly requiring a personal trainer.

Ahmed Pinto et al. [5] addressed the increasingly prevalent problem of obesity by creating a system to forecast how likely an obese patient is to reduce their weight as a function of their eating and exercise regimen. Employing machine learning techniques such as k-nearest neighbors (KNN) and decision trees, they created predictive models to assist medical practitioners in providing more tailored advice and guidance to patients. Their study also highlights how important it is to think about diet and exercise as a pair when developing successful weight control strategies.

Kulkarni et al. [6] were involved in designing personalized diet and exercise programs through the application of techniques like cosine similarity and Pearson correlation. They sought to aid patients suffering from chronic ailments like diabetes and hypertension by providing individualized lifestyle advice. Through their research, it is apparent how machine learning can be leveraged to help in patient care, particularly when managing long-term conditions.

Annapureddy et al. [7] emphasized the role of calorie control in a healthy lifestyle and demonstrated how artificial intelligence can be applied to provide customized diet and exercise recommendations. The uniqueness of their method is that their system considers every individual's needs and desires, allowing users to stick to healthy diets without having to give up on their favorite foods entirely.

Lastly, Sadhasivam et al. [8] constructed an integrated recommendation system that is personalized for diet and workout programs for every individual user. It collects essential personal details — such as height, weight, and age — and uses these details to determine the user's BMI. Using that, it builds tailored meal planning and exercise schemes meant to satisfy their own desired health outcomes. What sets their method apart is how it considers the larger picture — integrating calorie requirements, balance of nutrients, and fitness goals — to promote a balanced and sustainable strategy for overall health and wellness.

## III. PROPOSED METHODOLOGY:

This part describes the method employed in generating individualized workout schedules, detailing the process of classification as well as the methods applied to test the system's performance. Figure 1 presents a clear graphical illustration of the suggested methodology, demonstrating how the various components collaborate towards producing customized workout recommendations.

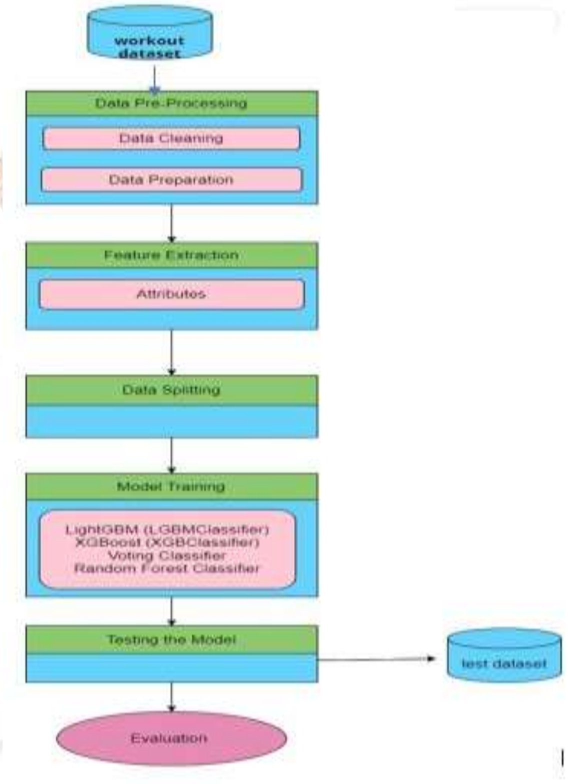


Figure 1 Architecture diagram

The data for the system is gathered from different sources of fitness exercises. The data set comprises important information like workout type, time, weight, set order, and reps. The aim is to make a prediction of the exercise name based on these inputs such that the system is able to recommend the most suitable exercise to every user based on their personal workout preferences.

**Workout Schedule Dataset Preprocessing:**

To normalize the Duration feature, which the users may input in varying formats (e.g., hours, minutes, or a mix of both), we used a custom function that normalizes all time data into a common total minutes format. This normalization makes all the inputs uniform and facilitates easier processing by the model. For features such as Workout Type, we used OneHotEncoding to transform them into machine-readable format. At the same time, numerical attributes such as Duration, Weight, and Set Order were input directly into the model with no further processing.

**1.1 Feature Extraction:**

The feature extraction for the workout schedule emphasizes major attributes required for personalization. Significant features are Workout Name, a categorical feature that is a representative of exercises such as squats or deadlifts, which is encoded using OneHotEncoding to be easily compatible with LightGBM. Numerical features like Duration (converted to minutes), Weight, Set Order, and Reps are directly provided to the model to enable personalization based on user-specific values. To ensure data quality, inconsistent time formats are made uniform, and missing values are replaced using the median or mean. Preprocessing in this manner provides a clean, trustworthy dataset, allowing LightGBM to produce correct workout schedules.

**2.Model Selection & Training**:

In this research, different machine learning models were evaluated to determine the best model that could produce tailored workout timetables. The chosen model was then trained with the dataset in order to look for patterns and enhance its prediction through a process of organized training and validation.

**1.LightGBM (LGBMClassifier):**

Algorithm: Gradient Boosting

We employed LightGBM for this system, which is a gradient boosting platform famous for being fast, efficient, and able to support big datasets. LightGBM creates a sequence of decision trees in which every subsequent tree identifies and fixes the mistakes made by its predecessor, enhancing the accuracy of overall prediction in stages. Among the main strengths of LightGBM is the capability to process categorical as well as numerical features without extensive preprocessing, so it's perfect for multi-variate fitness data. Prior to training, the data is passed through a

preprocessing stage to get all features into the correct format. Categorical features, such as Workout Type, are encoded, while numerical features, including Duration and Weight, are scaled where appropriate. LightGBM's emphasis on minimizing the loss function means that the model consistently improves itself, learning from its previous errors in order to make more accurate and tailored workout recommendations.

Due to its efficiency and capacity to manage intricate feature interactions, LightGBM assists the system in providing quick and accurate predictions, even when new data is brought into the mix.

**XGBoost (XGBClassifier) :**

Algorithm: Gradient Boosting

XGBoost is a high-performance and general gradient boosting library that is popular for its speed, accuracy, and capability to process enormous data. Just like LightGBM, XGBoost utilizes tree-based learners such that each tree attempts to fix the mistakes committed by the others, with an improvement in overall prediction accuracy with time. What is unique in XGBoost is that it has regularization techniques built-in, which avoid overfitting, and it's perfect for dense datasets that include both numeric and categorical features. It also offers parallel processing support, and it can thus learn faster on a number of CPU cores in parallel. XGBoost also provides support for user-defined objective functions, and that allows users to tune the model for a certain task. These characteristics render XGBoost a popular option within many types of machine learning tasks, including classification and regression, ranking and recommendation systems. Its universality and robustness render it an essential tool when creating personalized workout suggestions.

**Voting Classifier Model:**

Algorithm: Voting Ensemble Learning

Voting classifiers are ensemble techniques that make predictions based on the results of multiple models — in our case, models such as LightGBM and XGBoost. Rather than trusting a single model, a voting classifier makes predictions from all models involved and compiles the outcomes, generally by majority vote for classification tasks. This ensures a balance between the strengths and limitations of separate models, resulting in generally superior overall accuracy and more reliable predictions. The biggest benefit of employing voting classifiers is that they can eliminate the risk of over-dependency on the assumptions or weaknesses of a single model. By aggregating various models, each learned from slightly different views of the data, the system can learn more diverse sets of patterns and relationships, making it more robust. This renders voting classifiers particularly valuable for applications such as personalized exercise suggestion, where user interests and physical condition data can be extremely diverse.

**Random Forest Classifier**:

Algorithm: Ensemble Learning (Random Forest)

Random Forest Classifier is an extremely popular ensemble learning algorithm, which builds lots of decision trees, each of which has been trained on a slightly different subset of the data. When making a prediction — i.e., whether or not to recommend a workout — the model combines the output of all the individual trees to make a final prediction. The aggregation process minimizes overfitting and delivers overall accuracy and stability of predictions. One of the greatest strengths of Random Forest is its ability to handle categorical and numerical data, which makes it ideal for combining user preference, workout type, and performance measures. Since each tree has unique relationships and patterns in the data, combining them allows the model to discover more complex interactions — i.e., how each user's weight, workout level, and workout duration all factor into determining the best workout plan. By capitalizing on the variation among the trees, Random Forest gives personalized exercise recommendations that are suitable for one's needs so that people can have regimens that are suitable for their fitness goals.

**3. PERFORMANCE EVALUATION METRICS:**

1. Accuracy **-** Accuracy measures the model's overall performance in exercising name prediction against user input features (Workout Type, Duration, Weight). It is the ratio of accurate predictions to total predictions made by the model

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

- Where
  - $TP$ = True Positives
  - $TN$ = True Negatives
  - $FP$ = False Positives
  - $FN$ = False Negatives

2. Precision -Precision is a measure of how well the model's positive predictions are — in this instance, how many of the exercises the model suggested were indeed correct. It is calculated by determining the number of correctly suggested exercises (true positives) divided by the total number of exercises the model suggested as appropriate. High precision indicates that the model does not suggest irrelevant or incorrect workouts, so users receive more accurate recommendations.

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

3. Recall **-** Recall measures how accurately the model identifies all the right exercises for a particular user. It's the proportion of properly recommended exercises (true positives) out of all the exercises that in fact should have been recommended. High recall indicates that the model is less likely to leave out workouts that are suitable for the user, and thus is particularly critical in the process of making sure users receive all the alternatives that are suitable for their needs and fitness interests.

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

4. F1 Score - The F1 Score is a harmonic measure that averages precision (how many of the exercises recommended were correct) and recall (how many correct exercises were correctly identified). It's the harmonic mean of the two, so it's a good measure when both accuracy and completeness are equally valuable. In the context of this AI fitness coach system, a robust F1 Score not only ensures that the model suggests appropriate workouts but also captures all appropriate exercises that are aligned with the user's interests and fitness level — ultimately suggesting a balanced and customized workout plan.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Confusion Matrix - The confusion matrix provides a clear and comprehensive summary of how well the model is doing. It reveals both the number of good predictions (when the model was correct and when it correctly rejected an exercise), as well as the mistakes — instances where the model suggested the wrong exercise or failed to suggest a relevant one. This matrix allows us to know where the model is performing well and where it may require fixing, thus being an important tool in adjusting the workout recommendation system.

## IV. RESULTS AND DISCUSSION:

The LightGBM classifier performed the best among the models tried, with an accuracy of 89% and precision of 89%. It easily accommodated both categorical features such as workout names (using OneHotEncoding) and numerical features such as duration and weight, making it a good candidate for generating personalized workout schedules. While the model generally performed well, there were minor accuracy drops for some specific exercises. In spite of this, LightGBM was a stable and effective option for creating personalized workout routines.

One other popular gradient boost model, XGBoost, achieved 80% accuracy. Although XGBoost is perhaps best known as a versatile algorithm with excellent general performance across machine learning tasks generally, it could not match that of LightGBM in the current project. Although both of these models utilized tree learners as their foundation, differences in handling data and optimally training to achieve may have resulted in failure by XGBoost to perform as well as LightGBM. While XGBoost is still a strong and consistent model, LightGBM had better performance than it in acquiring the complex user input-workout recommendation interaction on this specific exercise app.

The Voting Classifier, which averaged the predictions of LightGBM and XGBoost together with a majority voting method, had an accuracy of 79%. In some cases, ensemble learning can improve performance by combining the strengths of many models, but not in this instance. In this project, the combined predictions didn't better LightGBM standing alone. This could be due to the fact that both base models were very comparable, i.e., they probably brought overlapping knowledge but not complementary strengths. This finding stresses the significance of selecting disparate models while constructing ensemble systems in order to really reap the advantages of having multiple strategies combined.

The Random Forest model performed at 69%, which was the weakest among all the tested models. While it is popular for performing well even on complex data and preventing overfitting through bagging predictions from various decision trees, Random Forest found it difficult to identify the subtle patterns between prominent workout features such as Workout Name, Duration, and Weight. Though the model's robustness against noisy data was evident, its constituent trees couldn't compete with the accuracy and flexibility of newer gradient boosting algorithms such as LightGBM and XGBoost. This outcome indicates that increased fine-tuning or refinement of the feature set might improve its performance in the case of generating personalized workouts.

| | Predicted Positive | Predicted Negative |
|---|---|---|
| Actual Positive | TP | FN |
| Actual Negative | FP | TN |

Figure 3 is a graph that compares the accuracy percentages of the various algorithms employed to create customized workout schedules. The graph shows the results for LightGBM, XGBoost, the Voting Classifier, and the Random Forest Classifier, providing a visual comparison of how each model fared. The side-by-side comparison serves to bring out which algorithms performed better in customizing workout suggestions based on user inputs.

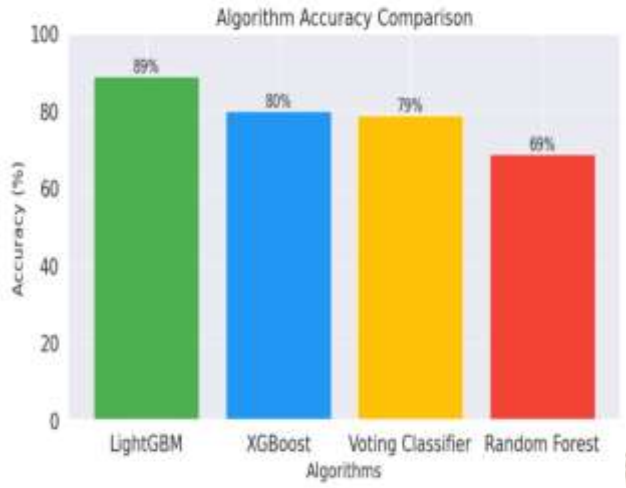


Figure 3 algorithm accuracy comparison

## V. CONCLUSION:

In conclusion, LightGBM proved to be the best-performing model with an accuracy of 89%, rendering it the most optimal option for creating customized workout timetables. Using the gradient boosting technique, in conjunction with decision trees, it was able to efficiently process categorical and numerical data, producing very high accuracy and tailored exercise suggestions. Its excellent performance makes it even more appropriate for use in applications that demand customized workout plans based on each user's preference, type of workout, duration, and weight.

Whereas XGBoost and the Voting Classifier produced decent outputs, their accuracy did not quite match that of LightGBM. Future development of the system might include further development of the LightGBM model, particularly through the addition of more user-specific data like fitness level, age, and individual fitness goals. More hyperparameter tuning and investigation of more varied ensemble techniques might improve accuracy and make the model even more universally applicable to a variety of users.

In the future, real-time feedback from users can be added to make the system more responsive and interactive so that it can fine-tune suggestions based on actual user experiences continuously. Creating a mobile app version, incorporating AI-driven progress monitoring, and incorporating the system in wearable devices would also enhance user engagement and develop a more extensive and customized fitness experience. Besides this, investigating future-proof deep learning methods may aid the system to identify more complex exercise patterns as well, hence improving its effectiveness.

## VI. REFERENCE:

1.Yadav, A., Choudhary, M., Gautam, P., Sankhe, K., & Sirsat, A. (2024). Workout Prediction Using Machine Learning. IEEE.

2.Felix, J., Mathu, T., & Roshni Thanka, M. (2023). LSTM Based Next Word Prediction for Workout Descriptions in an Android Mobile Application. IEEE.

3. Rehman, A., & Kumar, S. (2023). Leveraging MediaPipe and YOLO Keypoint Detection in Ensemble Approaches for Workout Pose Recognition. IEEE.

4. Rahman, A., Saji, A., Teresa, A., Nair, D. R., & Saritha, S. (2024). Physical Exercise Classification from Body Keypoints Using Machine Learning Techniques. IEEE.

5. Ahmed Pinto, K., Abdullah, N. L., & Keikhosrokiani, P. (2023). Diet & Exercise Classification using Machine Learning to Predict Obese Patient's Weight Loss. IEEE.

6. Kulkarni, K. S., Bansal, D., Bhagat, V. V., & Ganie, A. H. (2023). SmartHealth: Personalized Diet and Exercise Plans Using Similarity Modeling. IEEE.

7. Annapureddy, V., Keerthana, & Kumar, M. H. (2023). Diet and Exercise Tracking for Daily Requirements Using Artificial Intelligence Technique. IEEE.

8. Sadhasivam, S., Sarvesvaran, M. S., Prasanth, P., & Latha, L. (2023). Diet and Workout Recommendation Using ML. IEEE.

9. K. Lakshmi Joshitha, P. Madhanraj, B. Rithick Roshan,
G. Prakash, and V. S. Monish Ram, “AI-FIT COACH - Revolutionizing Personal Fitness With Pose Detection, Correction and Smart Guidance,” *IEEE*, 2023.

10. R. Pawar, B. Dhake, C. Chavan, U. Gohil, and R.
Ransing, “AI FIT Virtual Gym With Real-Time Exercise Form Detection,” *IEEE*, 2023.

11. A. Bilal Asghar, M. Majeed, A. Taseer, M. Burhan Khan, K. Naveed, and M. H. Jaffery, “Comparative Performance Analysis of Machine Learning Algorithms for Arm and Shoulder Exercises Using Wrist-Worn Band,” *IEEE*, 2023.

12. A. Smiley and J. Finkelstein, “Exercise Exertion Levels Prediction,” *IEEE*, 2023.

13. X. Min, L. Zeyu, S. Zheng, and L. Xunzi, “Prediction of College Students' Physical Test Effect Based on BP Neural Network,” *IEEE*, 2022.

14. S. R. Bakos and M. Sipos, “Analyzing Training Exercises with Mathematical Algorithms and Convolutional Neural Network,” *IEEE*, 2023.

15. H. Ma, Z. Huang, W. Tang, and X. Zhang, “Exercise Recommendation Based on Cognitive Diagnosis and Neutrosophic Set,” *IEEE*, 2022.

16. S. R. Bakos and M. Sipos, “Analyzing Training Exercises with Mathematical Algorithms and Convolutional Neural Network,” *IEEE*, 2023.