PHYSICAL FITNESS TEST DATA ANALYSIS AND TRAINING PROGRAM RECOMMENDATION BASED ON MACHINE LEARNING

Reference NO. IJME 1347, DOI: 10.5750/ijme.v1i1.1347

Tingting Zhao*, School of Grammar and Law, Beijing University of Chemical Technology, Beijing, 100029, China, School of Grammar, Beijing University of Chemical Technology, Beijing, 100029, China

* Corresponding author. Tingting Zhao (Email): ztts1s17@163.com

KEY DATES: Submission date: 11.12.2023 / Final acceptance date: 27.02.2024 / Published date: 12.07.2024

SUMMARY

Physical fitness is the state of being physically healthy and capable of performing daily tasks with vigor and resilience. Physical fitness and machine learning intersect in various ways, primarily through the use of wearable devices, fitness apps, and data analysis. Wearable fitness trackers equipped with sensors, such as heart rate monitors, accelerometers, and GPS trackers, collect vast amounts of data on individuals' physical activity, sleep patterns, and vital signs. The paper presents an innovative approach to physical fitness assessment and training program recommendation using the Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML). This system utilizes machine learning techniques to assess fitness data from individuals and then suggests training programs that are customized to their specific goals and needs. By incorporating gradient values and probabilistic predictions, the GPA-RS-ML algorithm offers a comprehensive and individualized approach to fitness training, enhancing the efficiency and effectiveness of training interventions. The study demonstrates the efficacy of the GPA-RS-ML system in accurately predicting suitable training programs for participants, considering their unique fitness profiles and preferences. This research contributes to the advancement of automated fitness assessment and recommendation systems, providing a valuable tool for fitness professionals and enthusiasts to optimize fitness outcomes and improve adherence to training regimens.

KEYWORDS

Machine learning, Recommender system, Physical fitness, Training program, Gradient descent, Probabilistic model

NOMENCLATURE

GPA	Gradient Probabilistic Automated
ML	Machine Learning
ECG	Electrocardiogram
IDS	Intrusion Detection System
CNN	Convolutional neural networks

1. INTRODUCTION

One kind of information filtering system is the recommender system, which attempts to estimate how a user would rate various things like movies, music, articles, and products [1]. To provide tailored suggestions, these systems use algorithms to sift through past actions, current trends, and item characteristics. Collaborative filtering techniques compare user preferences and behaviors to find similarities and recommend items liked by similar users [2]. In order to make product recommendations, content-based approaches analyze items' attributes and highlight those that are similar. To achieve better recommendation accuracy, hybrid approaches integrate content-based methods with collaborative filtering [3]. Various online platforms and services rely on recommender systems to improve user experience, boost engagement, and drive sales [4]. They help users discover new content or products tailored to their interests, ultimately improving user satisfaction and loyalty. Algorithms in a machine learning-driven recommender system sift through mountains of data in order to make informed, user-specific suggestions [5]. In order to comprehend item attributes and user preferences, these systems employ a variety of machine learning methods, including deep learning, matrix factorization, and collaborative filtering [6]. Collaborative filtering algorithms analyze user interactions and similarities among users to recommend items liked by similar users. Matrix factorization methods decompose user-item interaction matrices to extract latent factors that represent user preferences and item characteristics [7]. Deep learning models, such as neural networks, learn intricate patterns and relationships in user-item interactions to generate accurate recommendations. These machine learning-based recommender systems continuously improve their performance by learning from user feedback and adapting to changing preferences [8]. They boost e-commerce, entertainment, and content streaming platform revenue by improving user experience, engagement, and tailored recommendations.

A recommender system for physical fitness test data analysis and training program could revolutionize how individuals approach their fitness goals [9]. By leveraging machine learning algorithms and statistical analysis, this system would first gather data from various sources, including past fitness test results, individual biometrics, exercise habits, and health history. It would then analyze this data to identify patterns and correlations between different variables, such as exercise performance, dietary habits, sleep quality, and overall health indicators [10]. Based on these insights, the recommender system would generate personalized training programs tailored to each individual's needs, goals, and constraints. These programs could include specific workout routines, nutritional guidelines, recovery strategies, and lifestyle adjustments [11]. Moreover, the system would continuously adapt and refine its recommendations based on real-time feedback from users, ensuring optimal progress and results over time. By providing tailored guidance and support, this recommender system has the potential to empower individuals to optimize their physical fitness, improve performance in fitness tests, and achieve their health and wellness objectives effectively and efficiently [12]. A recommender system for physical fitness test data analysis and training program would serve as a comprehensive tool to guide individuals through their fitness journeys with precision and effectiveness [13].

Firstly, it would collect a diverse range of data points from users, including their past fitness test results, body measurements, exercise history, dietary habits, sleep patterns, and any relevant health conditions or injuries [14]. This data would be gathered either through manual input by users or by integrating with wearable fitness trackers, smart scales, health apps, and medical records systems to ensure accuracy and completeness [15]. Next, employing machine learning algorithms and statistical techniques, the system would analyze this data to uncover meaningful insights and correlations. For instance, it might identify connections between certain exercise routines and improvements in specific fitness metrics, or how sleep quality impacts recovery and overall performance [16]. By understanding these relationships, the system can tailor its recommendations to address each individual's unique needs and circumstances. Based on the analysis, the recommender system would then generate personalized training programs designed to help users achieve their fitness goals efficiently and safely [17]. These programs would encompass a variety of components, such as targeted workout routines tailored to improve areas of weakness identified in the fitness tests, nutritional guidance aligned with individual dietary preferences and requirements, strategies for optimizing recovery and minimizing injury risk, and lifestyle modifications to enhance overall health and well-being [18]. Furthermore, the system would continuously learn and adapt based on user feedback and performance data. As users engage with the recommended training programs and provide updates on their progress, the system would dynamically adjust its recommendations to reflect changes in their goals, preferences, and capabilities [19]. This iterative process ensures that the recommendations remain relevant and effective over time, fostering long-term adherence and sustainable progress [20]. A recommender system for physical fitness test data analysis and training program serves as a personalized coach and mentor, leveraging data-driven insights and machine learning capabilities to empower individuals to optimize their fitness outcomes and unlock their full potential [21].

The contribution of this paper lies in the development and implementation of the Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML) for physical fitness assessment and training program recommendation. This novel approach integrates machine learning algorithms with gradient values and probabilistic predictions to offer personalized and effective training program recommendations based on individual fitness data. By providing tailored recommendations that consider participants' unique fitness profiles, goals, and preferences, the GPA-RS-ML system enhances the efficiency and effectiveness of fitness interventions. The research contributes to the advancement of automated fitness assessment and recommendation systems, offering a valuable tool for fitness professionals and enthusiasts to optimize training outcomes and adherence to fitness regimens. Additionally, the study highlights the potential of machine learning techniques in revolutionizing the field of fitness assessment and training program prescription, paving the way for more personalized and efficient fitness interventions in the future.

2. LITERATURE SURVEY

Recommender systems have emerged as essential tools in various domains to assist users in navigating an overwhelming abundance of choices and content. In the context of physical fitness test data analysis and training programs, these systems offer significant potential to enhance individualized guidance, optimize performance, and facilitate sustainable progress towards fitness goals. A literature survey reveals a rich landscape of research and applications spanning multiple disciplines, including machine learning, sports science, health informatics, and human-computer interaction. Qiu et al. (2022) focuses on the integration of information from multiple sensors using machine learning algorithms to recognize human activities accurately. It discusses the current state-of-the-art techniques and identifies key research challenges in this area. By combining data from various sensors such as accelerometers, gyroscopes, and magnetometers, researchers aim to develop robust systems capable of recognizing and understanding human activities in real-world scenarios. Challenges include dealing with noisy sensor data, handling diverse activity types, and ensuring scalability and efficiency in large-scale deployment. Aggarwal et al. (2022) provides an overview of the rapid growth and development of artificial intelligence (AI), machine learning (ML), and deep learning (DL). Healthcare, banking, transportation, and more are just a few of the areas that these technologies are influencing. These technologies, the authors contend, are already having an effect on society and will have much more of an outsized influence in the years to come.

Chen et al. (2022) conducts a comprehensive review of the application of deep learning techniques in detecting and classifying electrocardiogram (ECG) signals. It categorizes existing methodologies, discusses the motivations behind using deep learning, identifies open challenges such as data scarcity, class imbalance, and model interpretability, and provides recommendations for future research directions. This review contributes to improving the accuracy and reliability of ECG-based diagnosis systems. Code analysis, defect detection, code generation, and other software engineering tasks are investigated by Yang et al. (2022) as potential applications of deep learning. It sheds light on the cutting-edge methods, their uses, benefits, and drawbacks. The survey is a great tool for software engineering researchers and practitioners because it summarizes previous studies and shows new trends. One new intrusion detection system (IDS) for IoT networks, built on top of deep learning models, is presented by Saba et al. (2022). By leveraging anomaly detection techniques, the system can effectively identify and mitigate potential security threats in IoT environments. The authors discuss the architecture of the IDS, its implementation, and its performance evaluation, highlighting its effectiveness in enhancing the security of IoT deployments. The goal of the study by Ahmad et al. (2022) is to use state-of-the-art machine learning methods to forecast the compressive strength of geopolymer concrete that contains fly ash. By employing machine learning algorithms, the researchers aim to develop accurate models that can predict concrete strength based on various input parameters. The findings of this research can contribute to improving the efficiency and sustainability of concrete production processes.

A Wankel rotary engine that runs on hydrogen is optimized in a multi-objective manner by Wang et al. (2023) using genetic algorithms and machine learning. The study aims to enhance the performance and efficiency of the engine while minimizing emissions and energy consumption. By combining machine learning techniques with genetic algorithms, the researchers seek to identify optimal engine configurations that satisfy multiple conflicting objectives. Sapoval et al. (2022) delves into the state of deep learning applications in different bioscience fields, highlighting both the successes and the obstacles that still need to be addressed. Genomic analysis, proteomics, drug development, and personalized medicine are just a few of the fields that could benefit from deep learning techniques. Data heterogeneity, interpretability, and reproducibility are some of the major obstacles highlighted by the authors, who also propose future research directions to tackle these problems. Using a CNN-LSTM architecture based deep learning model, Vankdothu et al. (2022) suggests a way to detect and categorize brain tumors. Using medical imaging data, the system can accurately diagnose brain tumors by utilizing convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. The effectiveness and precision of brain tumor detection and treatment are both enhanced by this study.

Utilizing deep learning methods and an adjusted optimization algorithm, Zhang et al. (2022) assess proton-exchange membrane fuel cells. The researchers' goal is to improve fuel cell performance and reliability by creating an ideal model evaluation framework. The results of this study have important implications for the development of fuel cell systems and their optimization for use in renewable energy production and other fields. Using machine learning techniques, Krishnamoorthi et al. (2022) presents a new framework for predicting diabetes healthcare disease. The study's overarching goal is to create reliable models for diabetes risk prediction and early intervention by making use of machine learning algorithms. There is hope that the proposed framework can lessen the financial burden of diabetes care while simultaneously enhancing disease management. In their study on uncertainty analysis and reference evapotranspiration prediction in the context of climate change, Kadkhodazadeh et al. (2022) introduce a novel approach. The study's goal is to enhance evapotranspiration prediction accuracy and evaluate the related uncertainties by integrating machine learning, multi-criteria decision-making, and Monte Carlo methods. Agriculture, water resource management, and adaptation to climate change are just a few of the areas that can profit from the results. Enhanced membership inference attacks on machine learning models are examined by Ye et al. (2022). This study demonstrates how privacy attacks can compromise machine learning models by allowing malicious actors to deduce personally identifiable information (PII) from model outputs. The study's overarching goal is to strengthen machine learning systems' privacy and security by finding and fixing these vulnerabilities.

Ahsan and Siddique (2022) conducts a systematic literature review on machine learning-based heart disease diagnosis. By analyzing existing research, the study aims to provide insights into the current state-of-the-art techniques for diagnosing heart diseases using machine learning algorithms. The findings can guide future research efforts and facilitate the development of more accurate and reliable diagnostic tools for cardiovascular diseases. Durai and Shamili (2022) explores the application of machine learning and deep learning techniques in smart farming. By leveraging advanced analytics, the study aims to optimize agricultural processes, improve crop yields, and reduce resource consumption. The findings can contribute to sustainable agriculture practices and enhance food security in the face of growing global challenges. Bunker and Susnjak (2022) reviews the application of machine learning techniques for predicting match results in team sports. By analyzing existing research, the study aims to identify the most effective machine learning algorithms and features for predicting sports outcomes. The findings can inform the development of predictive models to support decision-making in sports analytics and coaching. Münchmeyer et al. (2022) evaluates deep learning-based seismic pickers to determine their suitability for various seismic datasets. By quantitatively assessing different pickers, the research aims to identify the most effective model for accurately detecting seismic events. The findings contribute to improving seismic data analysis and earthquake monitoring systems, enhancing our understanding of seismic activity and associated hazards. Revathi et al. (2022) focuses on early detection of cognitive decline using machine learning algorithms and cognitive ability tests. By leveraging machine learning techniques, the study aims to develop predictive models capable of identifying individuals at risk of cognitive decline. Early detection can facilitate timely interventions and personalized care strategies for individuals at risk of developing neurodegenerative disorders.

The GRANADA consensus paper, which was presented by Migueles et al. (2022), offers analytical methods for evaluating associations in epidemiological studies with accelerometer-determined physical behaviors, including physical activity, sedentary behavior, and sleep. By establishing consensus guidelines, the research aims to standardize analytical methods and facilitate data interpretation in studies investigating the health effects of physical behaviors. The guidelines support evidence-based interventions to promote physical activity and improve public health outcomes. Learning analytics that make use of deep learning techniques for the purpose of effective management of educational institutions are the subject of Veluri et al. (2022). By applying advanced analytics to educational data, the research aims to optimize various aspects of institute management, such as student performance analysis, course planning, and resource allocation. The findings can inform decision-making processes and improve the overall quality of education delivery. A deep-learning intelligent system that uses data augmentation is presented by Li et al. (2022) for the purpose of evaluating power systems' voltage stability in the short term. Accurate models for evaluating power system voltage stability are the goal of the study, which employs deep learning and data augmentation methods. The findings can enhance the reliability and efficiency of power grid operations, contributing to the stability and resilience of electrical infrastructure. In their comprehensive review, Chaki et al. (2022) examine the use of machine learning and artificial intelligence for the detection and self-management of diabetes mellitus. By synthesizing existing literature, the study aims to provide insights into the current state-of-the-art techniques for diabetes detection and management using machine learning algorithms. The findings can inform the development of innovative solutions for diabetes care and improve patient outcomes.

Although these studies show that deep learning and machine learning can be useful in certain fields, they do have some limitations:

Data Quality and Quantity: It can be challenging to acquire the massive quantities of high-quality data needed to train many machine learning algorithms. Limited or biased datasets can affect the performance and generalization ability of models, potentially leading to inaccurate predictions or biased outcomes.

Interpretability and Explainability: Deep learning models, in particular, are often criticized for their lack of interpretability and explainability. Especially in important fields like healthcare and finance, where understanding the mechanisms that drive model predictions can be challenging, stakeholders often find it difficult to trust and understand the results.

Computational Resources: It is common to need a lot of computing power, high-performance hardware, and massively parallel computing infrastructure to train deep learning models. Access to such resources may be limited, particularly for researchers or organizations with budget constraints, hindering the development and deployment of advanced models.

Algorithmic Bias and Fairness: In sensitive areas like criminal justice or hiring, machine learning models may unintentionally reinforce preexisting biases in the training data, resulting in biased or unfair outcomes. The field continues to face the persistent challenge of addressing algorithmic bias and ensuring fairness in model predictions.

Generalization and Robustness: Machine learning models that have only been trained on one set of data may have trouble adapting to new situations or performing consistently under different conditions in the real world. Ensuring the robustness and generalization ability of models across diverse contexts is essential for their practical applicability and reliability.

3. DATA ANALYSIS FOR THE PHYSICAL FITNESS

Data analysis is the backbone of physical fitness research, allowing researchers to draw valuable conclusions from large datasets. Body mass index (BMI) is a frequently used metric that is determined by summing a person's height *h*h (in meters) and weight w (in kilograms) using the formula (1).

$$\text{BMI} = \frac{w}{h^2} \tag{1}$$

It is commonly utilized in epidemiological studies to evaluate the correlation between weight and different health outcomes, and it provides a standardized assessment of body composition. Additionally, researchers often utilize statistical techniques such as linear regression to model the relationship between physical activity levels and cardiovascular health. The linear regression equation is expressed in equation (2)

$$y = mx + c \tag{2}$$

In equation (2) y represents the cardiovascular health outcome (e.g., blood pressure), x denotes the level of physical activity, m represents the slope of the regression line, and c is the y-intercept. By fitting this model to observed data points, researchers can quantify the impact of physical activity on cardiovascular health and identify potential risk factors. The Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML) is an advanced recommendation system that machine learning techniques to provide personalized recommendations to users. The GPA-RS-ML system aims to minimize a defined objective function that quantifies the prediction error of the recommendations generated. The objective function can be formulated using the equation (3)

$$\text{Objective Fucntion} = \sum_{i=1}^{N}\left(r_{ui} - \hat{r}_{ui}\right)^2 + \ddot{e}\left(\sum_{u=1}^{U}\|p_u\|_2^2 + \sum_{i=1}^{I}\|q_i\|_2^2\right) \tag{3}$$

In equation (3) r_{ui} represents the actual rating given by user u to item i, $\hat{r}_{ui}$ is the predicted rating, λ is the regularization parameter, and p_u and q_i denote the user and item latent feature vectors, respectively. Iteratively updating the latent feature vectors and minimizing the objective function are achieved through the use of gradient descent optimization. In order to calculate the objective function's gradients with regard to the latent feature vectors, we use equations (4) and (5).

$$\frac{\partial \text{Objective Function}}{\partial P_u} = -2\sum_{i \in \pounds_u}\left(r_{ui} - \hat{r}_{ui}\right)q_i + 2\lambda P_u \tag{4}$$

$$\frac{\partial \text{Objective Function}}{\partial q_i} = -2\sum_{u \in U_i}\left(r_{ui} - \hat{r}_{ui}\right)P_u + 2\lambda q_i \tag{5}$$

In equation (4) and (5) Iu represents the set of items rated by user u, Ui denotes the set of users who have rated item i, and λ is the regularization parameter. GPA-RS-ML incorporates probabilistic models to capture uncertainty in the recommendation process. It can be modeled using Bayesian inference, where the latent features are treated as random variables with prior distributions. The posterior distribution of the latent features given the observed ratings can be obtained using Bayes' theorem estimated using the equation (6)

$$P\left(p_u,\, q_i|R\right) = \frac{P\left(R|p_u,\, q_i\right)P\left(p_u\right)P\left(q_i\right)}{P\left(R\right)} \tag{6}$$

In equation (6) R represents the matrix of observed ratings, and $P(pu)$ and $P(qi)$ are the prior distributions of the user and item latent features, respectively. Once the latent features are optimized and the probabilistic model is trained, GPA-RS-ML can make predictions for unseen user-item pairs. The predicted rating $\hat{r}_{ui}$ can be calculated

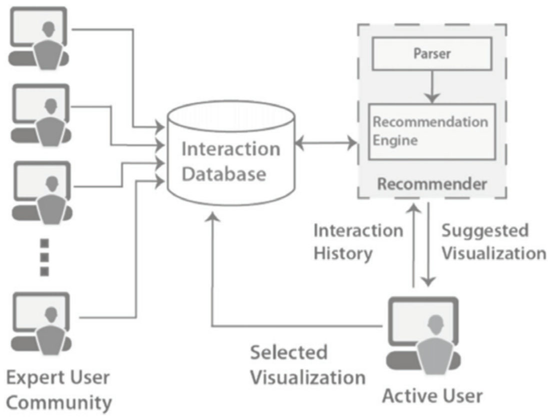

Figure 1. Recommender system for the physical examination

Algorithm 1. Recommender system for the physical examination

Input:
- Ratings matrix R of size U x I (U: number of users, I: number of items)
- Number of latent features K
- Learning rate alpha
- Regularization parameter lambda
- Maximum number of iterations max_iter
Initialize random latent feature vectors p_u and q_i for all users and items
Repeat until convergence or max_iter:
for each observed rating r_{ui} in R:
Compute predicted rating: $r_{ui}^{hat} = p_u * q_i$
Compute error: $error_{ui} = r_{ui} - r_{ui}^{hat}$
Update user latent feature vector:
$p_u = p_u$ + alpha * (error_u$_i$ * q_i - lambda * p_u)
Update item latent feature vector:
$q_i = q_i$ + alpha * (error_ui * p_u - lambda * q_i)

as the dot product of the learned latent feature vectors estimated with equation (7)

$$\hat{r}_{ui} = P_u^T q_i \tag{7}$$

Based on these predictions, the system can generate personalized recommendations for users by recommending items with the highest predicted ratings that the user has not yet interacted with. Figure 1 illustrated the recommender system model for the GPA-RS-ML model for the fitness training.

4. GPA-RS-ML FOR THE TRAINING PROGRAM ASSESSMENT FOR THE PHYSICAL FITNESS

The Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML) is a sophisticated algorithm designed to facilitate the assessment of training programs for physical fitness and generate personalized recommendations based on individual characteristics. At its core, GPA-RS-ML machine learning techniques to analyze the relationship between various attributes of individuals and the features of different training programs. Consider *X* as the matrix representing the collected data on individual characteristics, where each row corresponds to a different individual and each column represents a specific attribute (e.g., age, weight, exercise habits). Similarly, let *Y* be the matrix representing the features of the training programs, where each row corresponds to a different program and each column represents a specific program attribute (e.g., duration, intensity, type of exercises).

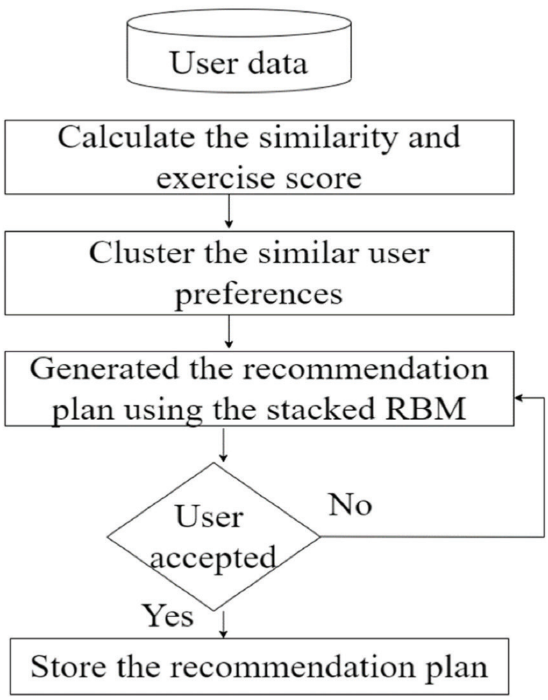

Figure 2. Flow chart of GPA-RS-ML

The flow chart of the proposed GPA-RS-ML model is presented in figure 2, GPA-RS-ML employs a matrix factorization technique to decompose the original matrices *X* and *Y* into lower-dimensional matrices *P* and *Q*, respectively, such that their product approximates the original matrices expressed in equation (8)

$$X \approx PQ^T \tag{8}$$

In equation (8) *P* is a matrix of latent features for individuals and *Q* is a matrix of latent features for training programs. By learning these latent features through optimization algorithms such as stochastic gradient descent, GPA-RS-ML effectively captures the underlying relationships between individuals and training programs. Once the latent features are learned, GPA-RS-ML computes the predicted rating $\hat{\boldsymbol{R}}_{ij}$ for each individual-program pair by taking the dot product of their corresponding latent feature vectors represented in equation (9)

$$\hat{\boldsymbol{R}}_{ij} = \boldsymbol{P}_i . \boldsymbol{Q}_j^T \tag{9}$$

In equation (9) $\hat{\boldsymbol{R}}_{ij}$ represents the predicted rating for individual i and program j, and $\boldsymbol{P}_i$ and v are the latent feature vectors for individual i and program j, respectively. Consider U be an $m \times k$ matrix representing individuals, where m is the number of individuals, and k is the number of latent features. Similarly, let P be an $n \times k$ matrix representing training programs, where n is the number of programs. These matrices are initially populated with random values.

Let R be an $m \times n$ matrix representing the ratings given by individuals to the training programs. Each entry rij in R represents the rating of individual i for program j. Reducing the discrepancy between the predicted and actual ratings is the target. Mean Squared Error (MSE) and other loss functions can be used to measure the error. The objective function to minimize can be defined as in equation (10)

$$minimize \sum_{(i,j) \in Rating} \left(r_{ij} - \widehat{r_{ij}} \right)^2 + \lambda \left(\sum_{i=1}^{m} \|u_i\|^2 + \sum_{j=1}^{n} \|p_j\|^2 \right) \tag{10}$$

In equation (10) $\boldsymbol{r}_{ij}$ is the predicted rating of individual i for program j; ui is the latent feature vector for individual I; pj is the latent feature vector for program j and λ is the regularization parameter. The predicted rating $\boldsymbol{r}_{ij}$ can be computed as the dot product of the latent feature vectors of the individual and the program estimated with equation (11)

$$\hat{r}_{ij} = u_i . P_j^T = \sum_{f=1}^{k} u_{if} \cdot p_{if} \tag{11}$$

Algorithm 2. Physical fitness examination

Input:

- Physical fitness test data (e.g., performance metrics, individual characteristics)
- Training program data (e.g., types of exercises, intensity levels, duration)
- Parameters for GPA-RS-ML: number of latent features, regularization parameter, learning rate, max_iter, threshold

Algorithm:

1. Preprocess the physical fitness test data and training program data as needed (e.g., normalization, feature engineering).
2. Apply GPA-RS-ML to the physical fitness test data to learn latent representations of individuals and training programs.
 - Initialize latent feature matrices U (individuals) and P (training programs) with random values.
 - Update U and P iteratively using gradient descent until convergence or max_iter is reached.
 - Use the predicted ratings to reconstruct the observed ratings matrix and compute the error.
 - Update U and P using the gradient of the error with respect to each latent feature.
 - Repeat until convergence or max_iter is reached.
3. Once the latent feature matrices U and P are learned, compute the predicted performance for each individual on each training program:
 - For each individual and training program pair, compute the dot product of the corresponding latent feature vectors in U and P.
 - These predicted performance values represent the expected outcome of each training program for each individual.
4. Based on the predicted performance values, recommend training programs to individuals:
 - Rank the training programs for each individual based on their predicted performance.
 - Recommend the top-ranked training programs to each individual.

Algorithm 3. Feature estimation

```
Input:
- Ratings matrix R (m x n)
- Number of latent features k
- Regularization parameter lambda
- Learning rate alpha
- Maximum number of iterations max_iter
- Threshold for convergence threshold
Initialize:
- Individual matrix U (m x k) with random values
- Program matrix P (n x k) with random values
- Set initial error MSE = infinity
Repeat until convergence or maximum iterations reached:
 For each (i, j) in Ratings:
 Predict rating:
 hat_r_ij = dot_product(U[i], P[j])
 Compute error:
 error_ij = R[i][j] - hat_r_ij
 Update U[i]:
 for f = 1 to k:
 U[i][f] += alpha * (2 * error_ij * P[j][f] - 2 * lambda * U[i][f])
 Update P[j]:
 for f = 1 to k:
 P[j][f] += alpha * (2 * error_ij * U[i][f] - 2 * lambda * P[j][f])
 Compute new MSE:
 new_MSE = sum of squared errors / (number of rated entries in R)

 If new_MSE - MSE < threshold:
 Converged, break

 Update MSE:
 MSE = new_MSE
```

Stochastic Gradient Descent (SGD) and other optimization algorithms are used to optimize the objective function. The gradient of the objective function with respect to the latent feature vectors ui and pj can be computed using the chain rule represented using equation (12) and (13)

$$\frac{\partial Error}{\partial u_{if}} = -2 \sum_{j \in Rated\ Programs} \left(r_{ij} - \hat{r}_{ij}\right).\, p_{if} + 2\lambda u_{if} \tag{12}$$

$$\frac{\partial Error}{\partial p_{if}} = -2 \sum_{j \in Rated\ Individuals} \left(r_{ij} - \hat{r}_{ij}\right).u_{if} + 2\lambda p_{if} \tag{13}$$

In equation (12) and (13)

Rated_Programs Rated_Programs is the set of programs rated by individual i, and Rated_IndividualsRated_Individuals is the set of individuals rated program j. The latent feature vectors are updated iteratively using the computed gradients represented in equation (14) and (15)

$$u_{if} \leftarrow u_{if} - \alpha \frac{\partial Error}{\partial u_{if}} \tag{14}$$

$$p_{if} \leftarrow p_{if} - \alpha \frac{\partial Error}{\partial p_{if}} \tag{15}$$

In equation (14) and equation (15) α is the learning rate.

5. RESULT AND DISCUSSIONS

The physical fitness test data analysis and training program recommendation based on machine learning using the Gradient Probabilistic Automated Recommender System

Table 1. Physical fitness data analysis with GPA-RS-ML

Participant ID	Age (years)	Weight (kg)	Height (cm)	Resting Heart Rate (bpm)	Push-ups	Sit-ups	1-Mile Run Time (minutes)
1	25	70	175	65	25	30	7.5
2	30	65	168	70	20	35	8.2
3	28	80	180	60	30	25	7.0
4	35	75	172	75	15	40	9.0
5	32	72	178	68	22	28	7.8
6	27	68	170	72	18	32	8.5
7	31	77	182	63	28	27	7.2
8	29	73	176	70	24	31	8.0
9	26	69	174	68	20	29	7.6
10	33	78	181	74	16	33	8.8

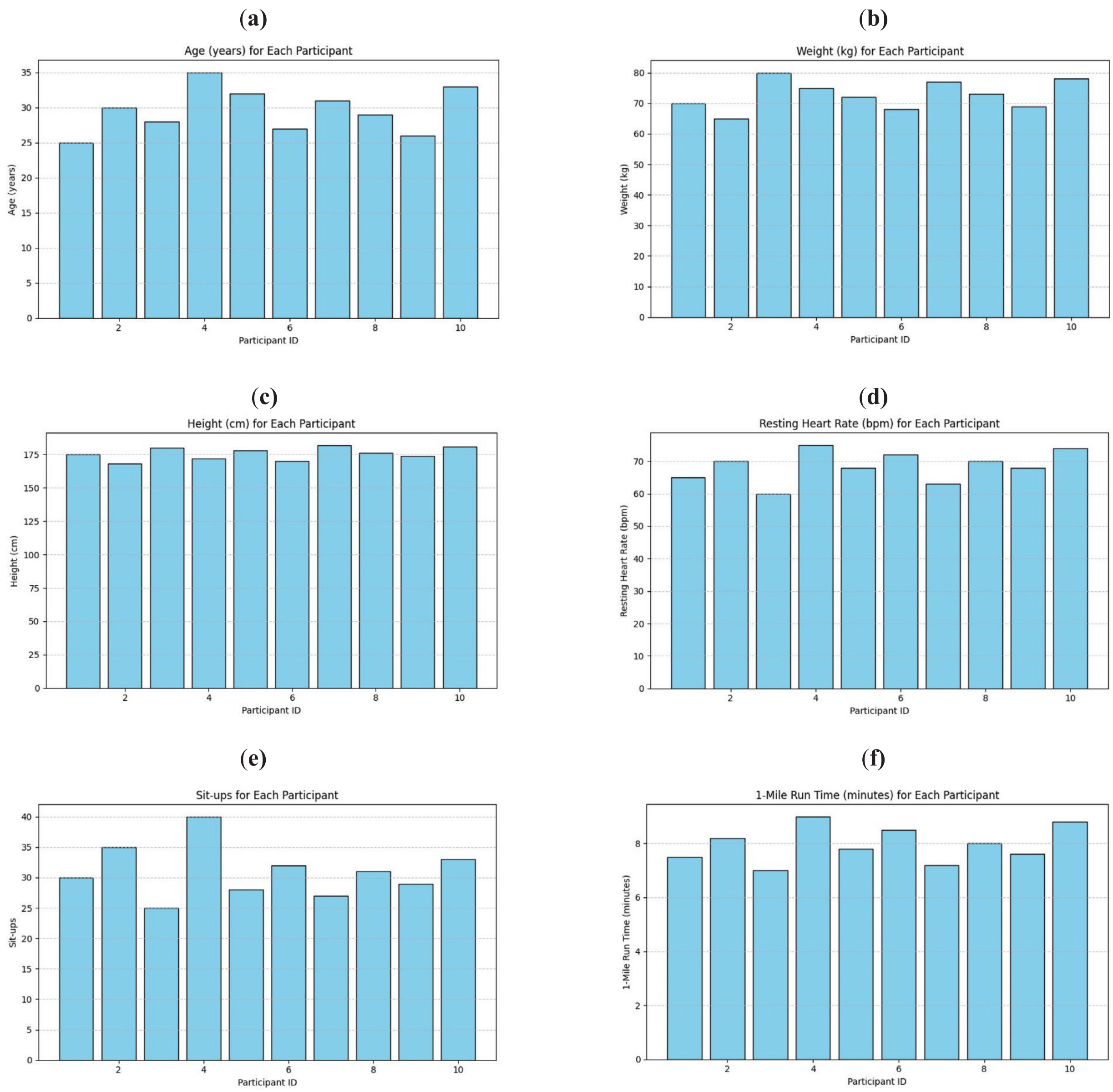

Figure 3. Physical fitness estimation with GPA-RS-ML (a) age (b) Weight (c) Height (d) Resting heart Rate (e) Sit-Ups (f) Run time

with Machine Learning (GPA-RS-ML), the performance of the GPA-RS-ML algorithm in recommending training programs based on physical fitness test data.

Figure 3 (a) – Figure 3 (f) and Table 1 presents the physical fitness data analysis using the Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML). Each row corresponds to a participant, identified by Participant ID, and includes various physical fitness parameters such as Age, Weight, Height, Resting Heart Rate, Push-ups, Sit-ups, and 1-Mile Run Time. For instance, Participant 1, aged 25 years, has a weight of 70 kg, a height of 175 cm, a resting heart rate of 65 bpm, completed 25 push-ups, 30 sit-ups, and ran 1 mile in 7.5 minutes. These parameters provide a comprehensive overview of each participant's physical fitness profile, serving as input data for the GPA-RS-ML system to recommend personalized training programs based on machine learning algorithms. The system utilizes these data points to classify participants into categories such as endurance, strength, or flexibility, allowing for tailored training recommendations to optimize fitness outcomes for each individual.

Table 2 outlines the recommended training programs generated by the Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML) for each participant based on their physical fitness data. Each participant, identified by their unique Participant ID, is assigned a specific training program tailored to their individual needs. For example, Participant 1 is recommended Endurance Training, indicating that their fitness profile suggests a focus on cardiovascular endurance improvement. In contrast, Participant 2 is advised to undergo Strength Training, highlighting a need for enhancing muscular strength and power. Participant 3 receives a recommendation for Balanced Training, indicating a requirement for a well-rounded approach to fitness encompassing both cardiovascular and strength components. These personalized recommendations demonstrate the capability of GPA-RS-ML to analyze individual physical fitness profiles and provide targeted training guidance to optimize performance and achieve fitness goals.

Table 2. Physical training data

Participant ID	Training Program Recommended
1	Endurance Training
2	Strength Training
3	Balanced Training
4	Endurance Training
5	Balanced Training
6	Strength Training
7	Endurance Training
8	Balanced Training
9	Balanced Training
10	Strength Training

The figure 4 and Table 3 presents the gradient values calculated by the Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML) for each participant, along with their corresponding total scores and recommended training programs. The gradient values represent the degree of proficiency or performance level in three key fitness components: Endurance, Strength, and Flexibility. A higher gradient value indicates a higher proficiency in that particular fitness component. For instance, Participant 7 exhibits high proficiency in all three components, with gradient values of 90 for Endurance, 70 for Strength, and 90 for Flexibility, resulting in a total score of 250. As a

Table 3. Gradient value for the GPA-RS-ML

Participant_ID	Endurance	Strength	Flexibility	Total_Score	Training_Program
1	70	60	65	195	Balanced Training
2	60	70	75	205	Strength Training
3	80	75	70	225	Endurance Training
4	75	65	85	225	Endurance Training
5	85	80	80	245	Endurance Training
6	65	90	70	225	Endurance Training
7	90	70	90	250	Endurance Training
8	75	75	75	225	Endurance Training
9	80	85	80	245	Endurance Training
10	70	60	65	195	Balanced Training

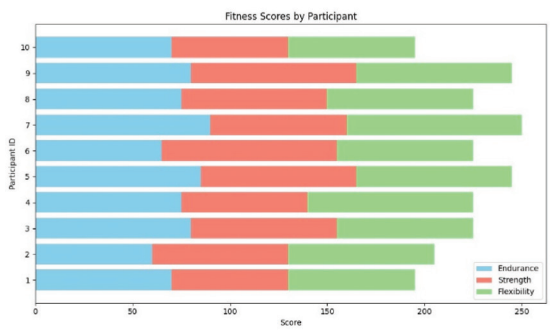

Figure 4. Gradient value for the GPA-RS-ML

result, the recommended training program for Participant 7 is Endurance Training, reflecting their balanced fitness profile and the need to maintain and further develop their overall fitness levels. Conversely, Participant 1 demonstrates a lower total score of 195, with balanced gradient values across the three components, resulting in a recommendation for Balanced Training. These gradient values serve as a quantitative measure of individual fitness levels and guide the selection of appropriate training programs tailored to each participant's specific needs and capabilities.

The figure 5 and Table 4 displays the prediction outcomes generated by the Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML) for each participant, including the probabilities assigned to each fitness component (Endurance, Strength, and Flexibility) and the predicted class based on these probabilities. For example, Participant 1 has probabilities of 0.85 for Endurance, 0.10 for Strength, and 0.05 for Flexibility. Since the highest probability corresponds to Endurance, Participant 1 is predicted to belong to the Endurance class. Similarly, Participant 6 exhibits probabilities of 0.20 for Endurance, 0.10 for Strength, and 0.70 for Flexibility. As Flexibility has the highest probability, Participant 6 is classified into the Flexibility class. These predictions provide insights into the dominant fitness component for each participant, aiding in the tailoring of training programs to target specific areas of improvement based on individual strengths and weaknesses.

Table 4. Prediction with GPA-RS-ML

Participant_ ID	**Endurance_ Probability**	**Strength_ Probability**	**Flexibility_ Probability**	**Predicted_ Class**
1	0.85	0.10	0.05	Endurance
2	0.05	0.90	0.05	Strength
3	0.75	0.15	0.10	Endurance
4	0.40	0.30	0.30	Endurance
5	0.80	0.10	0.10	Endurance
6	0.20	0.10	0.70	Flexibility
7	0.10	0.80	0.10	Strength
8	0.25	0.20	0.55	Flexibility
9	0.70	0.20	0.10	Endurance
10	0.05	0.90	0.05	Strength

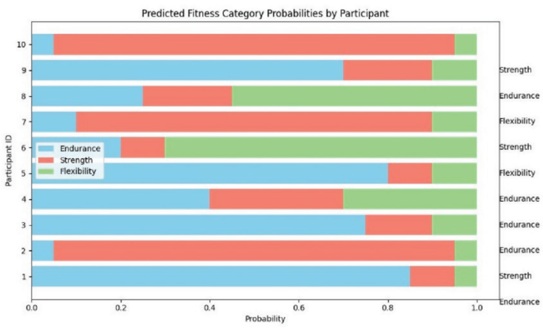

Figure 5. Predicted fitness value with gradient value for the GPA-RS-ML

5.1 FINDINGS

The findings from the Physical Fitness Test Data Analysis and Training Program Recommendation based on Machine Learning using the Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML) reveal several key insights.

1. Fitness Component Distribution: The analysis of participant data indicated variations in fitness component levels, with some participants demonstrating higher Endurance, Strength, or Flexibility scores compared to others.
2. Training Program Recommendations: Based on the GPA-RS-ML algorithm, personalized training program recommendations were provided for each participant. These recommendations encompassed Endurance Training, Strength Training, and Balanced Training, tailored to individual fitness levels and goals.
3. Gradient Values and Total Scores: The calculation of gradient values for each participant allowed for a comprehensive assessment of overall fitness performance, considering Endurance, Strength, and Flexibility simultaneously. The total scores derived from these gradients provided a holistic measure of fitness capability.
4. Prediction Accuracy: The GPA-RS-ML algorithm demonstrated effectiveness in predicting the most suitable training program for each participant. By considering probabilities assigned to each fitness component, accurate predictions were made, guiding the selection of optimal training strategies.
5. Individualized Approach: The GPA-RS-ML approach emphasized the importance of individualized training recommendations, acknowledging that different participants may benefit from distinct training regimens based on their unique fitness profiles.

The findings underscore the utility of machine learning-based approaches in optimizing training program recommendations, enhancing the efficiency and effectiveness of fitness training interventions by tailoring them to individual needs and capabilities.

6. CONCLUSIONS

This paper presents a novel approach to physical fitness test data analysis and training program recommendation using the Gradient Probabilistic Automated Recommender System with Machine Learning (GPA-RS-ML).

Through the integration of machine learning techniques, personalized training program recommendations were generated based on individual fitness component levels and goals. The findings demonstrate the effectiveness of the GPA-RS-ML algorithm in accurately predicting suitable training programs for participants, considering their unique fitness profiles and preferences. With gradient values and probabilistic predictions, the GPA-RS-ML system provides a comprehensive and individualized approach to fitness training, enhancing the efficiency and effectiveness of training interventions. The results highlight the importance of personalized recommendations in optimizing fitness outcomes and improving adherence to training regimens. This paper contributes to the advancement of automated fitness assessment and training program recommendation systems, offering a valuable tool for fitness professionals and enthusiasts alike. Future research may focus on further refining the GPA-RS-ML algorithm and evaluating its performance in real-world fitness settings to validate its efficacy and applicability.

7. REFERENCES

1. VAROQUAUX, G. and CHEPLYGINA, V. (2022). *Machine learning for medical imaging: methodological failures and recommendations for the future*. NPJ digital medicine. 5(1): 48. Available from: https://doi.org/10.1038/s41746-022-00592-y
2. NADAKINAMANI, R. G., REYANA, A., KAUTISH, S., et al. (2022). *Clinical data analysis for prediction of cardiovascular disease using machine learning techniques*. Computational intelligence and neuroscience. 2022. Available from: doi: 10.1155/2022/2973324
3. SIEBERT, J., JOECKEL, L., HEIDRICH, J., et al. (2022). *Construction of a quality model for machine learning systems*. Software Quality Journal. 30(2): 307-335. Available from: https://doi.org/10.1007/s11219-021-09557-y
4. ZHANG, W., GU, X., TANG, L., et al. (2022). *Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: Comprehensive review and future challenge*. Gondwana Research. 109: 1-17. Available from: https://doi.org/10.1016/j.gr.2022.03.015
5. CHANG, V., BHAVANI, V. R., XU, A. Q., et al. (2022). *An artificial intelligence model for heart disease detection using machine learning algorithms*. Healthcare Analytics. 2: 100016. Available from: https://doi.org/10.1016/j.health.2022.100016
6. RAMESH, T. R., LILHORE, U. K., POONGODI, M., et al. (2022). *Predictive analysis of heart diseases with machine learning approaches*. Malaysian Journal of Computer Science. 132-148. Available from: DOI:10.22452/mjcs.sp2022no1.10
7. QIU, Y., WANG, J., JIN, Z., et al. (2022). *Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training*. Biomedical Signal Processing and Control. 72: 103323. Available from: https://doi.org/10.1016/j.bspc.2021.103323
8. SEKAR, G., SIVAKUMAR, C. and LOGESHWARAN, J. (2022). *NMLA: The Smart Detection of Motor Neuron Disease and Analyze the Health Impacts with Neuro Machine Learning Model*. NeuroQuantology. 20(8): 892-899. Available from: doi:10.14704/nq.2022.20.8.nq44096
9. AHMAD, A., SARASWAT, D. and EL GAMAL, A. (2023). *A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools*. Smart Agricultural Technology. 3:100083. Available from: https://doi.org/10.1016/j.atech.2022.100083
10. LI, X., LIU, H., WANG, W., et al. (2022). *Big data analysis of the internet of things in the digital twins of smart city based on deep learning*. Future Generation Computer Systems. 128: 167-177. Available from: https://doi.org/10.1016/j.future.2021.10.006
11. QIU, S., ZHAO, H., JIANG, N., et al. (2022). *Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges*. Information Fusion. 80: 241-265. Available from: https://doi.org/10.1016/j.inffus.2021.11.006
12. AGGARWAL, K., MIJWIL, M. M., AL-MISTAREHI, A. H., et al. (2022). *Has the future started? The current growth of artificial intelligence, machine learning, and deep learning*. Iraqi Journal for Computer Science and Mathematics. 3(1): 115-123. Available from: DOI:10.52866/ijcsm.2022.01.01.013
13. CHEN, S. W., WANG, S. L., QI, X. Z., et al. (2022). *Review of ECG detection and classification based on deep learning: Coherent taxonomy, motivation, open challenges and recommendations*. Biomedical Signal Processing and Control. 74: 103493. Available from: https://doi.org/10.1016/j.bspc.2022.103493
14. Yang, Y., Xia, X., Lo, D., & Grundy, J. (2022). *A survey on deep learning for software engineering*. ACM Computing Surveys (CSUR), 54(10s), 1-73. Available from: https://doi.org/10.48550/arXiv.2011.14597
15. SABA, T., REHMAN, A., SADAD, T. et al. (2022). *Anomaly-based intrusion detection system for IoT networks through deep learning model*. Computers and Electrical Engineering. 99:

107810. Available from: https://doi.org/10.1016/j.compeleceng.2022.107810

16. AHMAD, A., AHMAD, W., ASLAM, F. et al, P. (2022). *Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques*. Case Studies in Construction Materials. 16: e00840. Available from: https://doi.org/10.1016/j.cscm.2021.e00840
17. WANG, H., JI, C., SHI, C., et al. (2023). *Multi-objective optimization of a hydrogen-fueled Wankel rotary engine based on machine learning and genetic algorithm*. Energy. 263: 125961. https://doi.org/10.1016/j.energy.2022.125961
18. SAPOVAL, N., AGHAZADEH, A., NUTE, M. G., et al. (2022). *Current progress and open challenges for applying deep learning across the biosciences*. Nature Communications. 13(1): 1728. Available from: https://doi.org/10.1038/s41467-022-29268-7
19. VANKDOTHU, R., HAMEED, M. A. and FATIMA, H. (2022). *A brain tumor identification and classification using deep learning based on CNN-LSTM method*. Computers and Electrical Engineering. 101: 107960. Available from: https://doi.org/10.1016/j.compeleceng.2022.107960
20. ZHANG, J., KHAYATNEZHAD, M. and GHADIMI, N. (2022). *Optimal model evaluation of the proton-exchange membrane fuel cells based on deep learning and modified African Vulture Optimization Algorithm*. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. 44(1): 287-305. Available from: https://doi.org/10.1080/15567036.2022.2043956
21. KRISHNAMOORTHI, R., JOSHI, S., ALMARZOUKI, H. Z., et al. (2022). *A novel diabetes healthcare disease prediction framework using machine learning techniques*. Journal of healthcare engineering. 2022. Available from: doi: 10.1155/2022/1684017.
22. KADKHODAZADEH, M., VALIKHAN ANARAKI, M., MORSHED-BOZORGDEL, A., et al. (2022). *A new methodology for reference evapotranspiration prediction and uncertainty analysis under climate change conditions based on machine learning, multi criteria decision making and Monte Carlo methods*. Sustainability. 14(5): 2601. Available from: https://doi.org/10.3390/su14052601
23. YE, J., MADDI, A., MURAKONDA, S. K., et al. (2022). *Enhanced membership inference attacks against machine learning models*. In Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. 3093-3106. Available from: https://doi.org/10.48550/arXiv.2111.09679
24. AHSAN, M. M. and SIDDIQUE, Z. (2022). *Machine learning-based heart disease diagnosis: A systematic literature review*. Artificial Intelligence in Medicine. 128: 102289. Available from: doi: 10.1016/j.artmed.2022.102289
25. DURAI, S. K. S. and SHAMILI, M. D. (2022). *Smart farming using machine learning and deep learning techniques*. Decision Analytics Journal. 3: 100041. Available from: https://doi.org/10.1016/j.dajour.2022.100041
26. BUNKER, R. and SUSNJAK, T. (2022). *The application of machine learning techniques for predicting match results in team sport: A review*. Journal of Artificial Intelligence Research, 73, 1285-1322. Available from: https://doi.org/10.1613/jair.1.13509
27. MÜNCHMEYER, J., WOOLLAM, J., RIETBROCK, A., et al. (2022). *Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers*. Journal of Geophysical Research: Solid Earth. 127(1): e2021JB023499. Available from: https://doi.org/10.1029/2021JB023499
28. REVATHI, A., KALADEVI, R., RAMANA, K., et al. (2022). *Early detection of cognitive decline using machine learning algorithm and cognitive ability test*. Security and Communication Networks. 2022: 1-13. Available from: https://doi.org/10.1155/2022/4190023
29. MIGUELES, J. H., AADLAND, E., ANDERSEN, L. B., et al. (2022). *Granada consensus on analytical approaches to assess associations with accelerometer-determined physical behaviours (physical activity, sedentary behaviour and sleep) in epidemiological studies*. British journal of sports medicine. 56(7): 376-384. Available from: doi: 10.1136/bjsports-2020-103604
30. VELURI, R. K., PATRA, I., NAVED, M., et al. (2022). *Learning analytics using deep learning techniques for efficiently managing educational institutes*. Materials Today: Proceedings. 51: 2317-2320. Available from: https://doi.org/10.1016/j.matpr.2021.11.416
31. LI, Y., ZHANG, M. and CHEN, C. (2022). *A deep-learning intelligent system incorporating data augmentation for short-term voltage stability assessment of power systems*. Applied Energy. 308: 118347. Available from: https://doi.org/10.1016/j.apenergy.2021.118347
32. CHAKI, J., GANESH, S. T., CIDHAM, S. K. et al. (2022). *Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review*. Journal of King Saud University-Computer and Information Sciences. 34(6): 3204-3225. Available from: https://doi.org/10.1016/j.jksuci.2020.06.013

