Article

Feedback Collection and Nearest-Neighbor Profiling for Recommendation Systems in Healthcare Scenarios

João António [1], Ricardo Malheiro [1,2] and Sandra Jardim [3,*]

1   School of Technology and Management, Polytechnic Institute of Leiria, 2411-901 Leiria, Portugal;
     2210786@my.ipleiria.pt (J.A.); rsmal@dei.uc.pt (R.M.)
2   Centre for Informatics and Systems, University of Coimbra, Pinhal de Marrocos, 3030-290 Coimbra, Portugal
3   Smart Cities Research Center, Polytechnic Institute of Tomar, 2300-313 Tomar, Portugal
*   Correspondence: sandra.jardim@ipt.pt

Abstract: The rise in the dimension and complexity of information generated in the clinical field has motivated research on the automation of tasks in personalized healthcare. Recommendation systems are a filtering method that utilizes patterns and data relationships to generate items of interest for a particular user. In healthcare, these systems can be used to potentiate physical therapy by providing the user with specific exercises for rehabilitation, albeit facing issues pertaining to low accuracy in earlier iterations (cold-start) and a lack of gradual optimization. In this study, we propose a physical activity recommendation system that utilizes a K-nearest neighbor (KNN) sampling strategy and feedback collection modules to improve the adequacy of recommendations at different stages of a rehabilitation period when compared to traditional collaborative filtering (CF) or human-constrained methods. The results from a trial show significant improvements in the quality of initial recommendations, achieving 81.2% accuracy before optimization. Moreover, the introduction of short-term adjustments based on frequent player feedback can be an efficient manner of improving recommendation accuracy over time, achieving overall better convergence periods than those of human-based systems, topping at a measured 98.1% accuracy at K = 7 cycles.

Keywords: recommendation systems; collaborative filtering; nearest neighbors; physical activity; knowledge extraction; machine learning; feedback collection; rehabilitation

1. Introduction

Developing some form of physical activity is a basic need for human physical and mental well-being and is essential in combating a sedentary lifestyle and its harmful effects [1]. Around 80% of the population does not practice enough physical activity to meet the recommendations of the World Health Organization (WHO). Regular physical activity speeds up metabolism, allowing for more effective fat burning and, consequently, weight loss, helping prevent and combat obesity [2,3]. Regular activity prevents cardiovascular diseases [4], as it plays an important role in cardiac rehabilitation as a means of reducing the morbidity and mortality rates associated with cardiovascular disease [5]. It also helps control diabetes [6,7], high blood pressure [8,9], and cholesterol levels [10], as well as improving sleep quality and increasing the ability to concentrate, leading to better academic and professional performance [11].

Health recommendation systems can encourage and engage users in modifying their behavior by providing improved options and practical knowledge, which are derived from the analysis of user behavior [12]. According to the technical taxonomy recently proposed by Etemadi et al. [13], healthcare recommendation systems can be classified into collaborative, content-based, knowledge-based, context-based, and hybrid techniques.

When it comes to evaluating healthcare recommendation systems, about half of the studies focused exclusively on the performance of the algorithms using several metrics,

while others conducted comprehensive randomized controlled trials or studies in the wild to assess their impact on human health conditions [12]. This reality suggests a slow progression of health recommendation systems, where their maturity and validation are not yet sufficient for their applicability in a real context.

Among the several possible applications of healthcare recommendation systems, such as recommending nutritional diets [14–18] or medicinal drugs [19–22], recommending physical activities as a means of promoting well-being or integrating rehabilitation processes has raised interest—both in the technical-scientific community and in the professional medical health sector—as a tool that potentially promotes better lifestyle habits. Although modern recommendation systems excel at providing accurate suggestions, they face several challenges and limitations, including issues with scalability, cold-start problems, data sparsity, and reliability.

Despite requiring more computational time in the classification process, especially with large datasets, the K-nearest neighbors (KNNs) classification method is one of the most used machine learning (ML) approaches in classification problems. As a non-parametric method, its basic principle is to classify a given data point based on the most common classification among the data points closest to it. Its simplicity and effectiveness in capturing complex patterns without assuming an underlying data distribution have made it one of the most used techniques in different contexts.

Research in healthcare settings using the K-nearest neighbors classification method has demonstrated significant potential in improving diagnostic accuracy and patient outcomes, having been applied to various medical datasets, addressing different problems such as disease prediction [23,24], patient risk stratification [25,26], and personalized treatment recommendations [27,28]. The algorithm's ability to handle large multidimensional datasets makes it particularly suitable for healthcare applications where data complexity is high. However, challenges remain in ensuring data privacy and integrating the KNN classification method with other ML methods to improve their predictive capabilities [29]. Overall, the KNN algorithm continues to be a valuable tool in medical research, contributing to more informed and timely healthcare decisions.

In this study, a new physical activity recommendation system is proposed as an intermediate layer between patients' activity sessions. The system is capable of generalizing patient profiles using KNN analysis, thus predetermining an initial exercise configuration that adequately matches the user's physical capacities. The method is expanded with a second operation mechanism, which periodically collects explicit feedback from patients. By using a bidirectional weight system, the system can model feedback directly into the recommendation logic, adjusting some aspects of the activity to patients' needs and preferences. Therefore, the method presented herein aims to contribute to the state of the art in the sense of bridging some of the difficulties with physical activity recommendation systems. In particular, problems such as data sparsity and cold-start are addressed by making adjustments and design choices that frequent auxiliary data collection and improve the quality of recommendations in first iterations.

This paper is organized into sections as follows. Section 2 introduces existing work in the context of healthcare recommendation algorithms, particularly those focused on rehabilitation practices. Section 3 presents the proposed approach, with a focus on the technical aspects of the components that make up the system's functionality. In Section 4, the results from a practical trial of the proposed method are presented and discussed. Finally, Section 5 presents conclusive statements and introduces potential lines of future research.

2. Related Work

The employment of recommendation systems covers a wide range of techniques and problem areas. In the past two decades, the growth in both the quality and quantity of systems is notable, namely in application areas related to streaming services, but also in industries such as healthcare and education. Below, an initial analysis is made of existing

approaches for recommendation systems, further detailing those applied to healthcare topics and identifying the most popular techniques and limitations of such methods.

Collaborative filtering (CF) is one of the more common techniques in building recommendation systems. In these approaches, items are recommended based on similarities in traits or actions with past users [30]. Moreover, recommendation systems based on CF are often designed to act on implicit or explicit feedback from the users, enabling personalized experiences [31].

As early as 2006, Aberg [32] proposed a meal-planning recommendation system for the elderly population. The method utilizes a mix of weighted soft constraints and traditional hard constraints for modeling patient satisfaction, meal preparation difficulty, and dietary restrictions, showcasing theoretical usability in real-world scenarios, although several drawbacks in algorithm speed and user experience are pointed out by the author. In 2010, Freyne and Berkovsky [33] presented a study on the sensitivity of a recipe recommendation system for groups of people to changes in weighing models and data aggregation heuristics, proposing that the best group recommendation performance is achieved when user's interests are individualized and scaled to group-based models.

The most common implementation of CF in healthcare is related to recommendation systems for connecting patients and doctors. In 2015, Narducci et al. [34] introduced Health-Net, a social network with a built-in recommendation system for creating recommendations between a patient and the doctor or medical institution that can provide them with the best possible service by utilizing proprietary measures of similarity between patient profiles and suggesting a ranked list of available doctors for the given patient's profile. Similarly, Guo et al. [35] developed a recommendation system in which the recommended doctors are ranked by their impact on the patient's specific health topic. The method has been extensively tested in China and shown to outperform benchmarks, supporting the usage of professional achievements as a key factor in making good recommendations. Later, in 2018, Han et al. [36] hybridized a traditional recommendation system to present patients with high-compatibility family doctors. Their approach is based on historical data and proposes the addition of a measurement of trust, having achieved performance that was superior to baseline heuristic or CF approaches.

Improvements to clinical decision support are also achievable with CF methods, namely in the field of therapy prescriptions. Gräßer et al. [37] propose a therapy recommendation system that aims to suggest efficient, systematic therapies for skin condition treatment based on a similarity measure between the current consultation and a large set of recorded consultations.

Recommendation systems for physical rehabilitation are closely linked to this study, and despite being a relatively immature subfield of research, some works support their real-world implementation. An example is the work presented by Gmez-Portes et al. [38], who proposed the application of fuzzy logic to improve the quality of therapeutic monitoring and guidance, suggesting a recommendation model that enables the remote rehabilitation of stroke patients through interactive exercises.

Ishraque et al. [39] leveraged an AI-supported chatbot for interfacing users with a recommendation system for cardiac rehabilitation exercises. The initial evaluation suggests that allowing users to personalize their therapy plans may induce better adherence and efficacy, though a real evaluation of these effects is not yet available. In addition, for cardiac rehabilitation scenarios, Tuijn et al. [5] introduced a tailored recommendation tool designed to incrementally boost physical activity and circumvent a sluggish start. The authors' proposed system harnesses a random forest classification model, amalgamating both measured and self-reported data to furnish individualized suggestions for physical activity. Moreover, the use of explainable artificial intelligence enhances clarity and fosters confidence in the system. The study underscores the capacity of ML approaches in tailoring recommendations for physical activity, advocating for a reinforcement learning strategy to refine system personalization as time progresses.

Ferreto et al. [40] proposed a personalized physical activity recommendation system for hypertensive patients to improve their lifestyle and health outcomes. The system uses a user profile model and a recommendation system, which considers cardiovascular risk factors, physical activity history, and physical activity result indices. Through validation by specialist doctors, 75% of the system's recommendations were approved, indicating its potential to help hypertensive patients improve their health through regular physical activity. The study also emphasizes the importance of personalized recommendations in promoting the well-being of individuals with hypertension.

By intending to restore and exercise the respiratory system, Zaboleeva-Zotova et al. [41] proposed a system of recommending specific exercises that are selected by each user. In this system, video analyses of physical activities carried out by users are used, taking into account their respective experiences. The selection of exercises to recommend is obtained by applying ML models.

The deep learning (DL) model proposed by Li et al. [42] aims to predict exercise recommendations made by experts, as well as the results of their clinical evaluations. The model uses patients' gait kinematic data to train a convolutional neural network. Regarding model evaluation, the authors report an accuracy of 78% in recommending a specific exercise.

3. Methods

In this paper, we propose a hybrid approach for recommending physical activities to elderly players of a gaming platform that targets the rehabilitation of upper limbs. Traditional CF strategies are employed for the initial sampling of patient profiles and the generation of baseline recommendations, while a built-in explicit feedback recovery system gradually models patients' specific preferences for a more personalized approach in subsequent iterations.

3.1. Data Characterization

All data were collected as a result of a clinical trial, aiming to assess the usability of a handheld dynamometer as a controller for augmented reality (AR) games. Participants were asked to play several games and consent to the respective data to be collected and analyzed. Each activity is comprised of a set of parameters that dictates its inherent difficulty level (length of the activity, proposed targets for grip strength, grip cadence, and number of grip presses). After each activity, participants could optionally provide feedback in the form of a 1-to-5 rating for that particular game. Exclusively, elderly patients with indicative traits of frailty were asked to participate in the trial. Two data sources are considered for the current study. For ease of explanation, they will be referred to as patient data (*PD*) and activity data (*AD*) in the characterizations below. It should be noted that both data sources have been previously submitted to a cleaning process for the replacement of missing values with statistical modes and the removal of outliers. As such, the characteristics presented are respective to clean data.

3.1.1. Patient Data

Patient data (*PD*) has 860 patient profiles, containing rows about general health and the degree of affliction with risk factors for frailty. Personal information is also present in the data, though it has been omitted for this study. In Table 1, a detailed view of PD is provided. Attribute names, data types, observed ranges, and average values are presented.

In Table 1, highlighted attributes refer to frailty risk factors and are the key target of this study. Despite having access to a variety of demographic and well-being indicators, not all attributes are explicitly explored in this study. More precisely, a correlation analysis between the existing attributes and their impact on frailty factors has ruled out the following attributes from the experimental process of the study: age, gender, number of hospitalizations, sleep quality, smoking, alcohol consumption, medication count, and

birthplace. To the extent of the analysis carried out, the aforementioned attributes appear to have negligible impact on the incidence of each frailty factor.

Table 1. Patient data characterization.

Attribute	Data Type	Range	Average/Mode
frailty	object	[non-frail; pre-frail; frail]	Pre-frail
gender	object	[F; M]	F
age	int64	[64, . . . , 87]	74
hospitalization one year	int64	[0, . . . , 6]	0
vision	object	[sees well; sees moderately; sees poorly]	Sees well
audition	object	[hears well; hears moderately; hears poorly]	Hears well
weight loss	object	[yes; no]	No
exhaustion score	int64	[1; 2]	1
balance single	object	[<5 s; >5 s; test non realizable]	>5 s
grip strength abnormal	object	[yes; no]	No
low physical activity	object	[yes; no]	No
bmi score	float64	[18.88, . . . , 53.08]	30.67
sleep	object	[no sleep problem; occasional sleep problem; permanent sleep problem]	No sleep problem
depression total score	int64	[0, . . . , 12]	2
anxiety perception	float64	[0, . . . , 13.6]	4.71
pain perception	float64	[0, . . . , 12.8]	5.07
regular activity	object	[no; <2 h per week; >2 h and <5 h per week; >5 h per week]	>2 h and <5 h per week
smoking	object	[never smoked; stopped (6 months); current smoker]	Never smoked
alcohol units	float64	[0, . . . , 35]	1.65
medication count	int64	[0, . . . , 18]	6
birthplace	object	[PT; ES; BR; AO]	PT

3.1.2. Activity Data

Activity data (*AD*) has 3525 rows of recorded game sessions performed by the patients in *PD*. These rows correspond to every valid activity collected during the trial, providing information about a relatively extensive range of game parameter combinations, from patient behavior to the achieved scores. Similarly to PD, an overview of the available attributes and observed values is provided in Table 2.

Due to this data source being mostly numerical, it is of relevance to assert that the grip force values are measured in kilograms of force (Kgf), and the grip timing values refer to the percentage of grip instances pressed at the correct time. A session is determined by a series of proposed targets for the patient in relation to their actual achieved scores, thereby justifying the presence of negative final score values as under-performing instances.

Table 2. Activity data characterization.

Attribute	Data Type	Range	Average/Mode
game	object	[GV001R; GV002M; GV003L]	GV001R
therapeutic difficulty	int64	$[0, \dots, 10]$	6
target grip force	float64	$[17.0, \dots, 25.9]$	20.4
target grip timing	float64	$[35.1, \dots, 89.0]$	61.0
target instances	int64	$[10, \dots, 50]$	28
grip force score	float64	$[13.2, \dots, 29.1]$	21.7
grip timing score	float64	$[20.0, \dots, 91.98]$	68.1
grip instances hit	int64	$[20.0, \dots, 91.98]$	25
final score	float64	$[-38.6, \dots, 47.4]$	5.72
session target duration	int64	$[45; 60; 90]$	58
session real duration	int64	$[23, \dots, 90]$	56
satisfaction query	float64	$[1; 2; 3; 4; 5]$	4.1

3.2. Hybrid KNN Sampling

First, recommendations are based on the CF principle that similar users will have similar preferences. The task of finding similarly disposed users can be solved with a variety of methods, although K-nearest neighbors is the preferred solution for this task due to its insensitivity to the number of classes, compatibility with multi-label classification, ease-of-deployment, and performance-to-accuracy ratio for small- to medium-sized datasets. An implementation of KNN is used to model the input feature array (new patient information) using a number, K, of previously existing feature arrays. Feature arrays are uniform in size, always containing 10 numerical and six categorical values pertaining to various physical and psychological measurements and the presence of frailty risk factors. Due to the coexistence of both categorical and continuous attribute types in the data, a single distance metric could not be used. Instead, categorical variables are transformed into binary arrays through attribute indication (dummy variable creation), and a weighted average of two measures is adopted when calculating neighbors.

The similarity score between numerical arrays A and B is given by the Euclidean similarity formula given by Equation (1). In order to avoid scale unbalance, the continuous variables in the data were scaled with a mean-based scaling formula, where the scaled value for a value, X_i, is given by $Xs = \frac{X_i - X_{mean}}{X_{sd}}$, where X_{mean} is the average value for the attribute and X_{sd} is the standard deviation for the attribute.

$$S_{A,B} = \frac{1}{1 + d(A,B)} , \tag{1}$$

where d is the Euclidean distance between both vectors, given by Equation (2).

$$d_{A,B} = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2} \tag{2}$$

Contrarily, the binary arrays corresponding to categorical variables in the data are evaluated with an adapted *Jaccard* similarity coefficient, as shown in Equation (3).

$$J_{A,B} = \frac{a}{a + b + c} , \tag{3}$$

whereby *a* is the count of attributes equal to 1 for both objects *A* and *B*; *b* is the count of attributes equal to 0 for object *A* and 1 for object *B*; and *c* is the count of attributes equal to 1 for object *A* and 0 for object *B*.

Finally, a weighing strategy was implemented to counter unfair similarity scores originating from having a larger number of continuous variables than categorical values (10 and 6, respectively). The corresponding weights were calculated as the relative fraction of total attributes per data type, or $10/16 = 0.625$ for continuous and $6/16 = 0.375$ for categorical data. As an example, if the Euclidean similarity between two patient rows is 0.95 but the Jaccard similarity is 0.50, the final similarity is given by $(0.95 \times 0.625) + (0.50 \times 0.375) = 0.78125$ or approximately 78.1%.

3.2.1. Determining *K*

The selection of an adequate number, *K*, of neighbors for generating classifications is often an overlooked decision when developing neighbor-based heuristic approaches. Selecting a sample size too small makes the system much less resistant to noise and other data quality inconsistencies, whereas an exaggerated *K* value is counterintuitive to the original premise of KNN itself. In order to find the best possible value for this problem, *K* was tested for all integer values in the range $[5, 86]$, with the upper limit corresponding to 10% of the dataset size. The metric used for this test pertains to the classification accuracy of the KNN algorithm in predicting previously assigned labels with a K-Means clustering algorithm, where $K = 4$. The results can be observed in Figure 1.
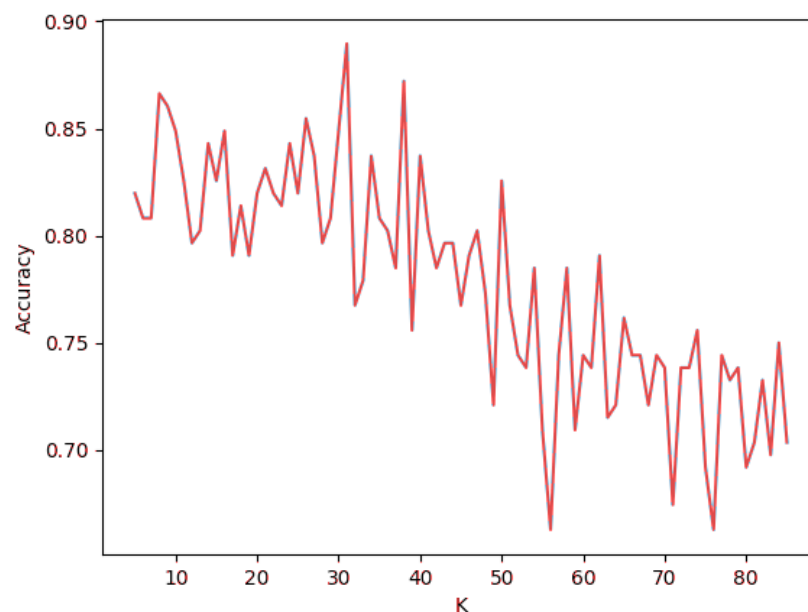

Figure 1. KNN classification accuracy for $K \in [5, 86]$.

The graph depicted in Figure 1 highlights an irregular behavior in which the accuracy value has pronounced variations for every step. This is credited to having used new samples of data for every value of *K*. In particular, a 20% split of the data (172 rows) was built on every iteration. This practice exposes the model to varying levels of neighbor influence, thus eliminating bias from single-test scenarios. Despite not following a predictable pattern, the highest observed classification accuracy for this range is 88.9% and relates to $K = 31$. Moreover, accuracy clearly tends to be lower as *K* increases beyond the 45–50 range. This region corresponds to when neighbors present bigger distances, thus defeating the purpose of KNN in finding densely packed points of data. This value will be used for the purpose of this study and the foreseeable scope of the project. However, a reiteration of the process outlined above may be necessary in the event of a significant expansion in the size of the dataset.

3.2.2. Scalability

Despite its logical simplicity, KNN-based classification algorithms often suffer from scaling issues once the dataset size increases substantially. In part, this could be accredited to the brute-force mechanism used in calculating distances to all points in the training set when determining the final list of neighbors, which is an $\mathcal{O}(N)$. Although this factor can predetermine that the proposed method will suffer from scalability issues, it is important to understand how the depreciation of the system evolves with increases in the number of observations in the data and in the value of K. Following this, a scalability assessment is conducted for the classification process of the proposed method. When evaluating the degradation for increases in observations, a data augmentation strategy was employed to add many synthetic rows to the patient dataset. The classification for $N = 1$ samples as the number of rows, Nr, increases can be consulted in Figure 2.
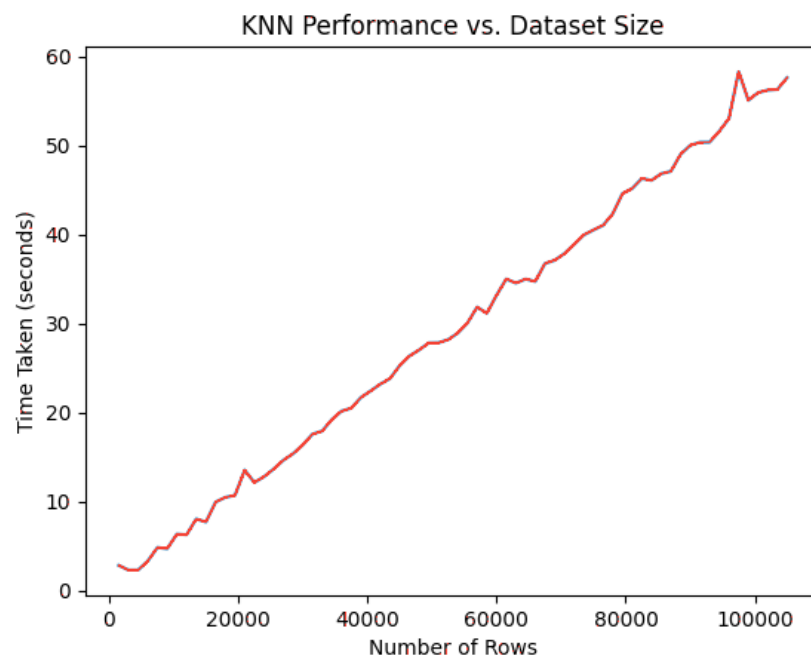

Figure 2. KNN classification times for $Nr \in [1000, 100{,}000]$.

The data in the graph are self-explanatory, as they model an $\mathcal{O}(N)$ algorithm predictably. This naturally indicates that the system will suffer from decaying performance as the number of patients used in KNN sampling increases. Another indicator of performance can be studied in the way classification time responds to increasingly higher K values for generating predictions. Decidedly, classification accuracy far outweighs classification time in selecting an optimal value for K, but the latter metric can provide an interesting secondary dimension to consider when future-proofing the system. Figure 3 contains measurements for $K \in [5, 86]$ using the standard 860 rows for classification.

In this regard, K does not have a significant effect on the classification time by itself. Despite observable spikes around $K = 30$, $K = 55$, and $K = 80$, no classification exceeded 1 s.

The information gathered with this portion of the study enables a deeper understanding of the viability of the KNN classification mechanism in the long term. Ultimately, the decision can be attributed to the promptness requirements of the recommendation system, the foreseeable growth of the data, and the accuracy trade-off with the employment of alternative strategies. As of presently, the classification accuracy achieved using KNN as a sampling method justifies its usage in favor of other, more performant methods, although future research could be directed into optimizing the sampling algorithm into an approximate nearest neighbors strategy instead, should the number of patients grow significantly.

Moreover, as recommendations are generated in the background and not on demand, near real-time responses are not a priority.
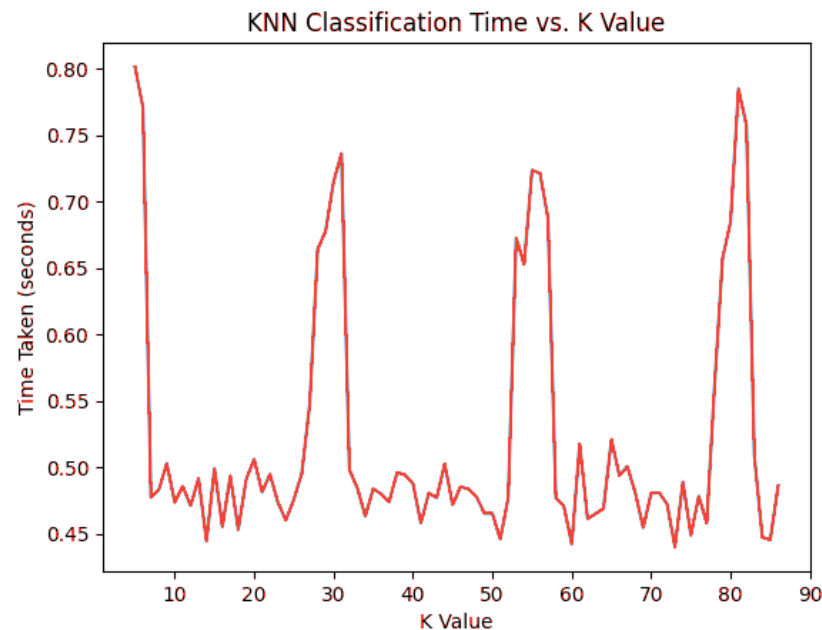

Figure 3. KNN classification times for $K \in [5, 86]$.

3.3. Generating Recommendations

The recommendations generated by the system are essentially a set of activity rules that should align with a patient's therapeutic needs, physical capacity, and, notably, their preferences. Two working modes were devised to generate recommendations, depending on whether the user has previously interacted with the system or not. Initially, implicit feedback is taken from the patient's overall health profile, whereby the activities played by a number of high-similarity neighbors are analyzed. Subsequent interactions with the recommendation system are built on explicit feedback and are collected at frequent intervals. The usage of both implicit and explicit feedback ensures that recommendations made with no prior knowledge of a patient can still be relevant, but this promotes a gradual change into a personalized experience as more feedback is collected.

3.3.1. Nearest-Neighbor-Based Recommendations

First, recommendations are generated in accordance with the general tendencies of the $K = 31$ nearest neighbors to that particular user. The workflow for instantiating the first recommendations can be consulted in Figure 4.

All neighbors' activity history is initially aggregated by rating, meaning that the game with the highest average satisfaction among the current demographic is identified first. In order to create a therapeutic experience that is cohesive with the user's physical capabilities, all individual difficulty levels for the game are tabulated in regards to their quality score, Qs, given by Equation (4), where the user's average engagement and satisfaction scores for that particular difficulty, x, are weighed. Given that the satisfaction score S takes values in $[0, 5]$, whereas the engagement score E varies in $[0, 1]$, the first is multiplied by a factor of 0.2 to avoid outweighing the latter by scale differential.

$$Qs_x = ((S_x \times 0.2) \times 0.5) + E_x \times 0.5 \tag{4}$$

Upon finding the difficulty level with the highest quality score, the specific targets for grip force, grip timing, and grip instances are generated randomly within the constraints defined for that difficulty.
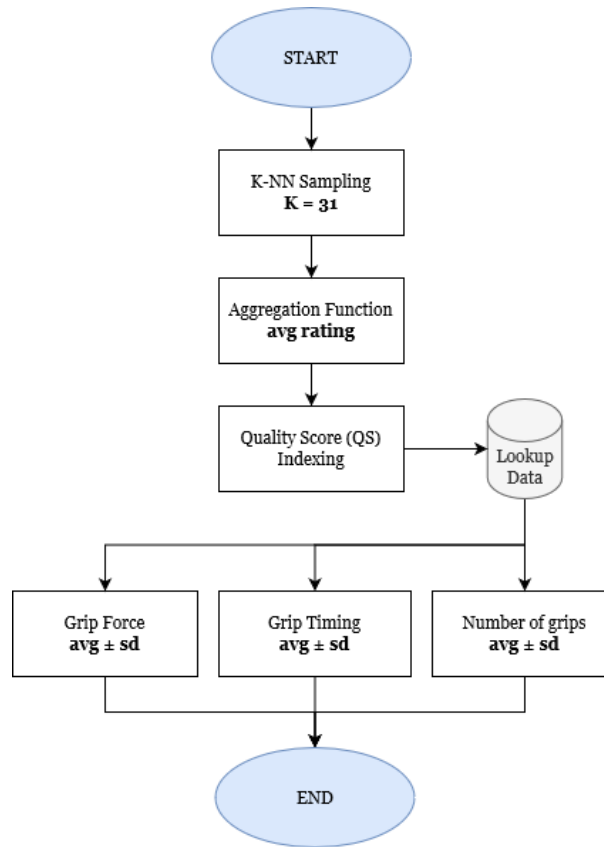
Figure 4. Workflow for $N = 1$ recommendations.

3.3.2. Weighed Feedback-Based Recommendations

After the first recommended activity is played, the recommendation workflow changes into a feedback-based approach based on the immediate impressions the users had about the activity they have just participated in. For this, a set of weights, *W*, is initialized for the current user. Weights are used to track the user's sentiment towards the specific components of the activity and are gradually adjusted through the application of a short questionnaire. The different questions, choice paths, and respective effects on the weights can be consulted in Figure 5. The questionnaires are designed with ease of use in mind, predicting an environment in which the majority of players are elderly and, thus, less accustomed to complex technology.

Any given weight set has a fixed number of positions, specifically 5, corresponding to the different activity elements that can be gradually fine-tuned to the user's preference. The weight-shifting mechanism can incur changes in the following:

- The game, either to avoid patient burn-out or to correspond to their preferences;
- The session duration, depending on whether the user found it to be excessive or insufficient. This makes the system responsive to the patient's energy levels and disposition to partake in longer activities;
- The session's grip targets, namely the recommended average grip force, the correct grip cadence (the interval at which grips are meant to be pressed), and the total number of times the patient needs to perform a grip. All of these parameters are initially bound by a predetermined difficulty level ranging from 0 to 10; however, to favor a personalized experience, individual targets are gradually optimized to the patient's specific needs. For example, a patient whose initial activity corresponds to a difficulty of 6 might possess a much steeper level of lethargy, requiring far fewer grips to account for their reduced energy, hence gradually having only that target adjusted accordingly.
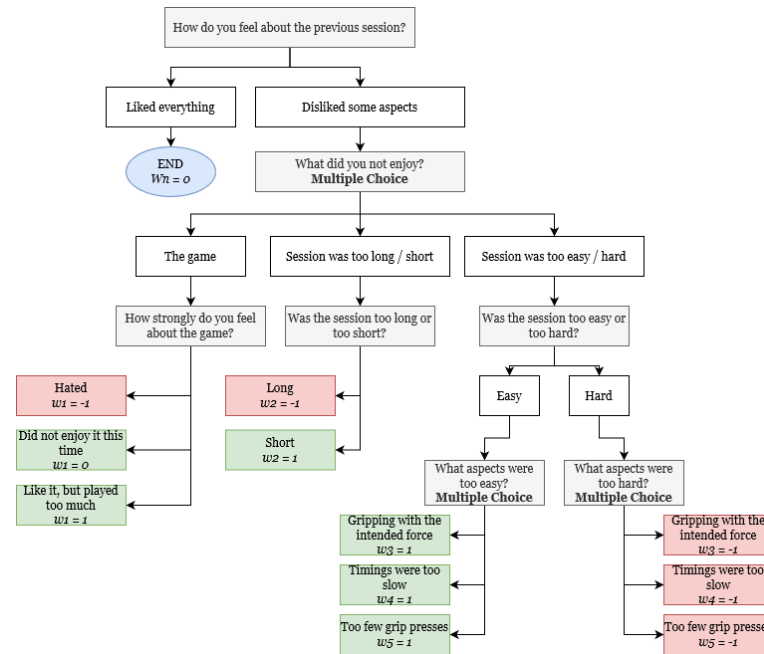
Figure 5. Feedback collection workflow, with respective bidirectional weight−shifting operations.

A bidirectional shift is used to inflict changes in weights, depending on whether the patient's sentiment towards a question is positive or negative. Generally, answers that indicate that the patient is struggling to achieve the proposed goals (negatively connoted) correspond to the respective weights being shifted to -1. Contrarily, if any goal is too easy to achieve (positively connoted), the corresponding weight is shifted to 1. Finally, when patients are satisfied with a particular setting, its weight is not altered from 0 (neutrally connoted).

3.3.3. Feedback Quality Assurance

Feedback provided by a system's users is subject to significant quality and quantity constraints, where many factors can play a role in precluding the quality of such data. In specific to the subject matter at present, feedback collected from elderly patients can be directly influenced by tiredness, lack of technical perception (when the patient cannot fully understand the purpose of feedback collection), unwillingness to co-operate, and malicious intent (when the patient purposely requests easier challenges to have perceivable improvements in scores). In order to help mitigate these effects, questionnaire application is normalized by a set of rules and supervision, namely the following:

- A "cool down" period is employed between activities, allowing patients to properly rest and not make decisions based on their state of tiredness;
- Activities with outlier scores for a given patient (abnormally high or low) are not considered for the questionnaire, as they do not accurately represent the patient in a "normal" state;
- The questions and the order in which they are presented are constant throughout every feedback collection process.
- An option to listen to the questions is provided to users with interpretation or reading difficulties.

Despite it being impossible to ensure non-malicious compliance in this scenario, the conditions put in place before feedback collection mitigate some of the more pressing issues when dealing with the elderly population.

4. Results

The proposed method was used by 71 returning patients from the original trial, enabling a comparative evaluation of the before and after states of the system. The trial was carried out under no supervision, with participants having full access to the recommendation system throughout its 3-day duration. At the end of the trial, the participants answered some questions about their overall experience. All participants played a total of 344 activities across the duration of the trial.

4.1. Model Assessment

Recommendation systems have the goal of returning items of interest to a given user. As such, the evaluation metrics for these systems typically differ from traditional systems, depending on their purpose. In the context of this work, the success of a recommendation is retrieved directly from the user. An activity is considered correct when the patient, after participation, it, does not request major changes to any of its difficulty settings. As such, accuracy is expected to grow as the number of iterations increases, which corresponds to the number of parameter adjustment cycles needed to achieve an adequate fit to the patient's needs. For this evaluation, the system is considered to converge when at least 97% of recommendations are considered correct (i.e., the accuracy of the system reaches 97%). In the previous (baseline) approach, recommendations are completely defined by human interaction. Therapists manually analyze recent activity data to find trends based on performance indicators such as patient scores and set up new recommendation parameters. With the goal of automating the analysis and recommendation set-up processes, the proposed system is hereby compared with the previous approach. The data shown in Figure 6 suggest that the proposed system is far superior in N_1 accuracy (i.e., first recommendations), supporting the usage of KNN for generalizing patient profiles from historical data.
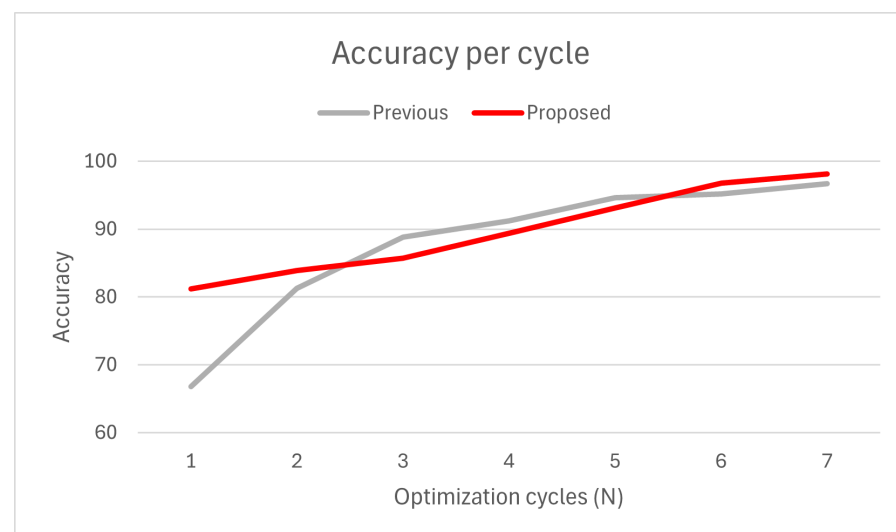

Figure 6. Average accuracy: Previous vs. proposed method.

Notably, the previous approach has a faster convergence speed in the earlier iterations, facing an increase of 22% accuracy from N_1 to N_3 and stabilizing from there onwards. Disregarding the N_1 iteration, the previous method has a slightly higher average convergence rate per cycle (2.56%) than the proposed method (2.36%). This is likely due to the previous method not being limited by a learning rate (and being human-bound), which also explains why the proposed method's accuracy increases steadily across iterations. Despite the slower convergence, the proposed method generally converges quicker ($N = 5$) than the previous method ($N = 6$) due to its superior N_1 accuracy.

Due to limitations in the scope and timeline of the trial, we were unable to determine the evolution of curves for $N > 7$, raising questions about the adequacy of the solution

for extended usage. Moreover, accuracy evaluation in recommendation systems does not account for situations that can not be modeled into data, such as the player's mood (the player may dislike activities they enjoyed in the past) and uncertainty (the player may not be sure about certain aspects of the activity), which will ultimately affect accuracy at irregular intervals. Follow-up studies should prioritize tackling questions of scalability and error to further solidify the position of the method in the state of the art.

4.2. Behavioral Evaluation

An indirect way of measuring the quality of the recommendations is by analyzing patients' behavior and adherence to the activities before and during the trial. Engagement is a metric that relates each activity's proposed duration with the actual time spent by the patient partaking in it and can indicate situations of early quitting due to exhaustion or boredom. For the purpose of the study, a patient is considered engaged with the activity when they participate in at least 95% of its duration, meaning that an implied goal of the system is to improve the ratio of players over this threshold.

Additionally, a patient's self-perception of improvement is a psychological indicator that the underlying system is working as intended. After participating in the trial, patients are asked to rate their performance subjectively on a scale from 0 to 100. The proposed approach was compared with the previous solution regarding the two behavioral metrics. First, the proportion of patients that match or exceed the proposed engagement score, and second, the average self-rating score. The results from this comparison can be seen in Figure 7.
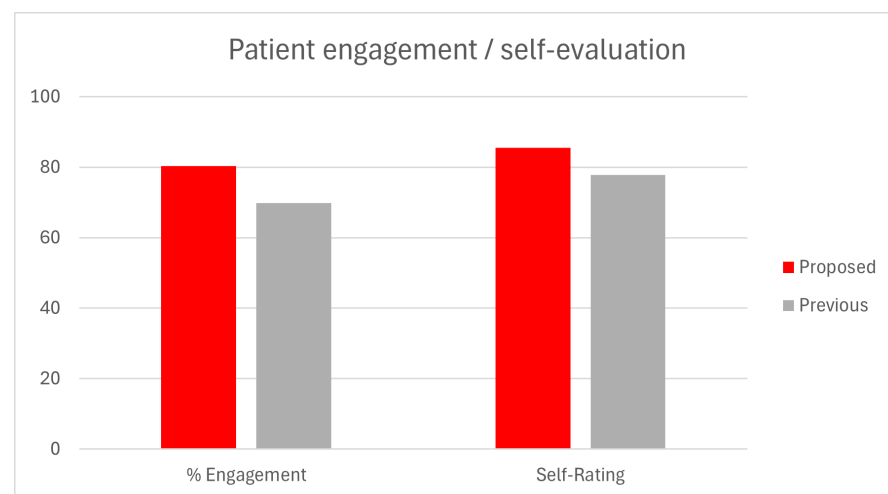

Figure 7. Behavioral indicators: Previous vs. proposed method.

The data depicted in the graph indicates a slight improvement in both indicators over the previous method. A 10.5% increase in the percentage of patients who achieved the proposed engagement scores is consistent with the uptick in first recommendation accuracy previously observed, meaning that when patients have initially suggested a better activity, they tend to have better adherence for the remainder of the rehabilitation plan. Patients also appear to rate their performances, on average, 8% higher than with the previous method, which could indicate a situation where activities are better suited to the patient's capabilities and preferences.

Feedback collected from patients after the trial revealed that a significant portion of players (66%) felt noticeable positive changes in difficulty between activity cycles. Generally, the bidirectional weight-shifting mechanism seems appropriate for instantiating adjustments according to the patient's specific needs. However, 11 participants claimed that the feedback questionnaire is too frequent and disruptive to a seamless experience, suggesting that an alternative optimization frequency might benefit user experience to the detriment of convergence speed.

4.3. Statistical Analysis

In order to support the claim that the proposed method improves upon the previous one, a statistical analysis was carried out. Due to both systems being comparable in the sense that only one class is possible for recommendations (when recommended equals true), the problem can be essentially considered a binary classification. Hence, statistical analysis consists of analyzing the observed proportions for each method. Both methods were evaluated regarding the average observed N_1 accuracy for the same set of users across a similar number of played activities. Figure 8 depicts the observed confidence intervals for both methods at N_1. The non-overlap between confidence intervals suggests that the proposed method has consistently better accuracy in the first recommendations.

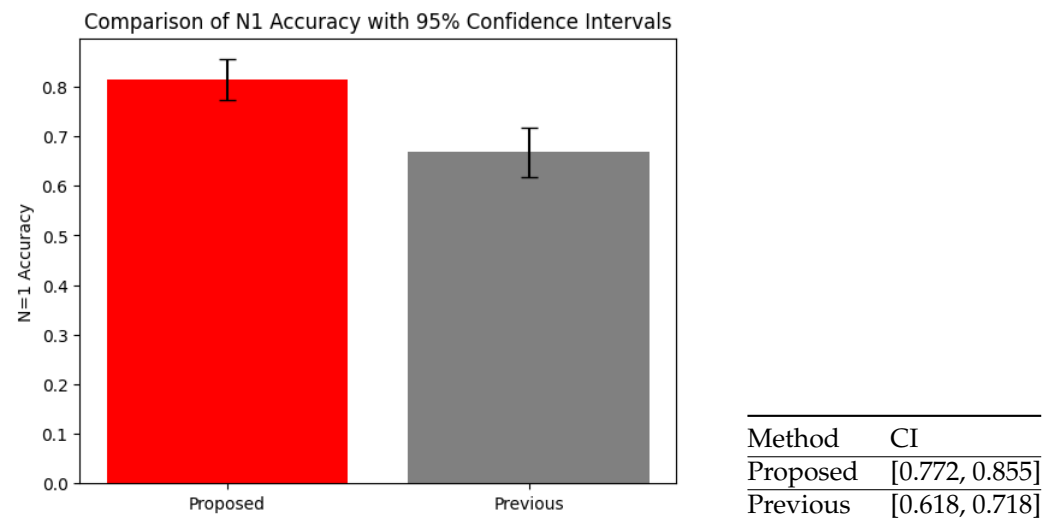| Method	CI
Proposed	[0.772, 0.855]
Previous	[0.618, 0.718]

Figure 8. Confidence intervals (CI): Previous vs. proposed method.

In order to further solidify the claim, a proportion Z-test was conducted under the null hypothesis that accuracy is the same in both methods ($\alpha = 0.05$). Given that the observed *p*-value equals 1.134×10^{-5}, a value less than 0.05, the null hypothesis can be rejected, which validates our statement that the proposed method is statistically significantly better than the previous one.

5. Conclusions and Future Work

Engaging in regular physical activity is proven to improve quality of life in several aspects, but despite the outlined benefits, a large portion of the population lacks sufficient exercise levels, leading to physical and psychological complications, especially in the latter stages of life. Physical therapy is a type of health rehabilitation practice that aims to reduce pain and improve motor function, promoting recovery from injury, medical intervention, and age-related loss of mobility. Recommendation systems play an important role in enabling personalized healthcare and, thus, can be leveraged to improve the efficacy of physical therapy through the autonomous prescription of rehabilitation exercises according to patients' specific needs.

The proposed method's functionality is defined by two distinct modes. First, a nearest neighbor-based collaborative filtering strategy is used to generate initial recommendations based on similarity in patient profiles. Then, explicit feedback collection interacts with a personalized weight system, gradually fitting newer recommendations to the patient's preferences. The weight-shifting mechanism is designed to provide the patient with an immediate and tangible change in the way their next activity plays out, enabling a degree of positive or negative adjustment to activity goals, depending on perceived difficulty.

The evaluation of the proposed method shows that there is a significant improvement in first recommendation accuracy when compared to a human-based approach, supporting

the role of multiple feature weighing for improving nearest neighbor analysis precision. Indicators of success can also be derived from behavioral data, which suggests that having high-quality recommendations in the earlier stages of the program can lead to better adherence to the activities. Consequently, when players' feedback has noticeable effects on recommendations, their understanding of self-improvement is improved. Overall, these results support the introduction of explicit feedback collection as a means of enhancing the personal aspect of recommendation systems applied to healthcare, in particular, physical activity recommendation systems. To the extent of our knowledge and research, the combination of strategies employed in this study is novel for the subject matter. As such, the present work expands the existing state-of-the-art healthcare recommendation systems on several fronts. More specifically, the study supports the mitigation of issues of model cold-start by using multiple combined feature-weighing employed in the context of neighbor analysis, as well as the employment of reinforcement learning methods based on explicit user feedback to improve a system's continuity and personalization.

As indicators for future work, it is necessary to conduct further research into the limitations of the feedback collection method, as well as evaluate the effects of the long-term usage of the system. In particular, it is essential to implement measures to prevent the exploitation of feedback collection. One such instance occurs when a patient knowingly and untruthfully reports their activities as being too difficult, thereby inducing changes that are not aligned with their actual physical capabilities. Moreover, a solution for preventing disruptive feedback collection needs to be explored, potentially involving the evaluation of alternative optimization periods to reduce the interruption of the player's experience. Finally, a subsequent study is planned to introduce the periodic inference of patients' performance. This will add another layer of feedback whereby patients can better understand their rehabilitation progress over time.

Author Contributions: Conceptualization, J.A.; Formal analysis, J.A.; Funding acquisition, S.J.; Investigation, J.A.; Methodology, J.A.; Project administration, R.M.; Software, J.A.; Supervision, R.M.; Validation, J.A. and R.M.; Writing and Editing J.A., S.J. and R.M. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Techframe's Human Sciences Ethic Committee, under the protocol code GV20231122TF, on 17 May 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy matters, and none of the data contained in the dataset mentioned in the manuscript are publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AD	Activity data
AR	Augmented reality
CF	Collaborative filtering
CI	Confidence intervals
DL	Deep learning
KNN	K-nearest neighbors
ML	Machine learning
PD	Patient data
WHO	World Health Organization

References

1. Galloway, M.T.; Jokl, P. Aging Successfully: The Importance of Physical Activity in Maintaining Health and Function. *J. Am. Acad. Orthop. Surg.* **2000**, *8*, 37–44. [CrossRef] [PubMed]
2. Burton, L.J.; VanHeest, J.L. The Importance of Physical Activity in Closing the Achievement Gap. *Quest* **2007**, *59*, 212–218. [CrossRef]
3. Steinbeck, K.S. The importance of physical activity in the prevention of overweight and obesity in childhood: A review and an opinion. *Obes. Rev.* **2008**, *2*, 117–130. [CrossRef] [PubMed]
4. Youngand, D.R.; Steinhardt, M.A. The Importance of Physical Fitness versus Physical Activity for Coronary Artery Disease Risk Factors: A Cross-Sectional Analysis. *Res. Q. Exerc. Sport* **2013**, *64*, 377–384. [CrossRef]
5. van Tuijn, R.; Lu, T.; Driesse, E.; Franken, K.; Gajane, P.; Barakova, E. WeHeart: A Personalized Recommendation Device for Physical Activity Encouragement and Preventing "Cold Start" in Cardiac Rehabilitation. *Lect. Notes Comput. Sci.* **2023**, *14144*, 191–201. [CrossRef]
6. Laaksonen, D.E.; Lindström, J.; Lakka, T.A.; Eriksson, J.G.; Niskanen, L.; Wikström, K.; Aunola, S.; Keinänen-Kiukaanniemi, S.; Laakso, M.; Valle, T.T.; et al. Physical Activity in the Prevention of Type 2 Diabetes: The Finnish Diabetes Prevention Study. *Diabetes* **2005**, *54*, 158–165. [CrossRef] [PubMed]
7. Shi, B.Y. The importance and strategy of diabetes prevention. *Chronic Dis. Transl. Med.* **2016**, *2*, 204–207. [CrossRef] [PubMed]
8. Park, S.; Han, K.; Lee, S.; Kim, Y.; Lee, Y.; Kang, M.W.; Park, S.; Kim, Y.C.; Han, S.S.; Lee, H.; et al. Cardiovascular or mortality risk of controlled hypertension and importance of physical activity. *Heart* **2021**, *17*, 1472–1479. [CrossRef]
9. Langhammer, B.; Bergland, A.; Rydwik, E. The Importance of Physical Activity Exercise among Older People. *BioMed Res. Int.* **2018**, *2018*, 7856823. [CrossRef]
10. Mann, S.; Beedie, C.; Jimenez, A. Differential Effects of Aerobic Exercise, Resistance Training and Combined Exercise Modalities on Cholesterol and the Lipid Profile: Review, Synthesis and Recommendations. *Sports Med.* **2014**, *44*, 211–221. [CrossRef]
11. Janssen, I.; Clarke, A.E.; Carson, V.; Chaput, J.P.; Giangregorio, L.M.; Kho, M.E.; Poitras, V.J.; Ross, R.; Saunders, T.J.; Ross-White, A.; et al. A systematic review of compositional data analysis studies examining associations between sleep, sedentary behaviour, and physical activity with health outcomes in adults. *Appl. Physiol. Nutr. Metab.* **2020**, *45*, S248–S257. [CrossRef]
12. Croon, R.D.; Houdt, L.V.; Htun, N.N.; Štiglic, G.; Abeele, V.V.; Verbert, K. Health Recommender Systems: Systematic Review. *J. Med. Internet Res.* **2021**, *23*, e18035. [CrossRef] [PubMed]
13. Etemadi, M.; Abkenar, S.B.; Ahmadzadeh, A.; Kashani, M.H.; Asghari, P.; Akbari, M.; Mahdipour, E. A systematic review of healthcare recommender systems: Open issues, challenges, and techniques. *Expert Syst. Appl.* **2023**, *213*, 118823. [CrossRef]
14. Zhang, J.; Wang, Z.; Liu, W.; Liu, X.; Zheng, Q. A unified approach to designing sequence-based personalized food recommendation systems: Tackling dynamic user behaviors. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 2903–2912. [CrossRef]
15. Rostami, M.; Aliannejadi, M.; Oussalah, M. Towards Health-Aware Fairness in Food Recipe Recommendation. In Proceedings of the 17th ACM Conference on Recommender Systems, Singapore, 18–22 September 2023; pp. 1184–1189. [CrossRef]
16. Kobayashi, A.; Mori, S.; Hashimoto, A.; Katsuragi, T.; Kawamura, T. Functional Food Knowledge Graph-based Recipe Recommendation System Focused on Lifestyle-Related Diseases. In Proceedings of the IEEE 18th International Conference on Semantic Computing, Laguna Hills, CA, USA, 5–7 February 2024; pp. 261–268. [CrossRef]
17. Brintha, N.C.; Nagaraj, P.; Tejasri, A.; Durga, B.V.; Teja, M.T.; Pavan, M.N.V. A Food Recommendation System for Predictive Diabetic Patients using ANN and CNN. In Proceedings of the 7th International Conference on Communication and Electronics Systems, Coimbatore, India, 22–24 June 2022; pp. 1364–1371. [CrossRef]
18. Thongsri, N.; Warintarawej, P.; Chotkaew, S.; Saetang, W. Implementation of a personalized food recommendation system based on collaborative filtering and knapsack method. *Int. J. Electr. Comput. Eng.* **2022**, *12*, 630–638. [CrossRef]
19. Suryadevara, C.K. Towards Personalized Healthcare—An Intelligent Medication Recommendation System. *Int. Eng. J. Res. Dev.* **2020**, *5*, 16.
20. Bhoi, S.; Lee, M.L.; Hsu, W.; Tan, N.C. REFINE: A Fine-Grained Medication Recommendation System Using Deep Learning and Personalized Drug Interaction Modeling. *Adv. Neural Inf. Process. Syst.* **2024**, *36*, 24013–24024.
21. Tan, W.Y.; Gao, Q.; Oei, R.W.; Hsu, W.; Lee, M.L.; Tan, N.C. Diabetes medication recommendation system using patient similarity analytics. *Sci. Rep.* **2022**, *12*, 20910. [CrossRef]
22. Yue, W.; Wang, M.; Zhang, L.; Zhang, L.; Huang, J.; Wan, J.; Xiong, N.; Vasilakos, A.V. A-GSTCN: An Augmented Graph Structural–Temporal Convolution Network for Medication Recommendation Based on Electronic Health Records. *Bioengineering* **2022**, *10*, 1241. [CrossRef]
23. Sateesh Kumar, R.; Sameen Fatima, S. Heart Disease Prediction Using Extended KNN (E-KNN). In *Smart Computing Techniques and Applications. Smart Innovation, Systems and Technologies*; Springer: Singapore, 2021; Volume 24. [CrossRef]
24. Shahade, M.; Awate, A.; Nandwalkar, B.; Kulkarni, M. Diabetes Disease Prediction Using KNN. In *Innovations in Data Analytics*; Springer: Singapore, 2023; Volume 1442. [CrossRef]
25. Borzooei, S.; Briganti, G.; Golparian, M.; Lechien, J.R.; Tarokhian, A. Machine learning for risk stratification of thyroid cancer patients: A 15-year cohort study. *Eur. Arch. Oto-Rhino-Laryngol.* **2024**, *281*, 2095–2104. [CrossRef]
26. Tripathi, R.; Khatri, S.; Van Greunen, D.; Ather, D. Risk Stratification of Breast Cancer Patients: Integrating Epidemiology, Risk Factors, and Prognostic Markers for Sustainable Development. In *Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2023; Volume 1939. [CrossRef]

27. Wu, Y.; Zhang, L.; Bhatti, U.; Huang, M. Interpretable Machine Learning for Personalized Medical Recommendations: A LIME-Based Approach. *Diagnostics* **2023**, *13*, 2681. [CrossRef] [PubMed]

28. Hu, Y.; Huerta, J.; Cordella, N.; Mishuris, R.G.; Paschalidis, I.C. Personalized hypertension treatment recommendations by a data-driven model. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 44. [CrossRef]

29. Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.* **2020**, *5*, 227–234. [CrossRef] [PubMed]

30. Pazzani, M.J.; Billsus, D. Content-based recommendation systems. In *The Adaptive Web: Methods and Strategies of Web Personalization*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 325–341.

31. Hu, Y.; Koren, Y.; Volinsky, C. Collaborative filtering for implicit feedback datasets. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 263–272.

32. Aberg, J. Dealing with Malnutrition: A Meal Planning System for Elderly. In Proceedings of the AAAI Spring Symposium: Argumentation for Consumers of Healthcare, Stanford, CA, USA, 27–29 March 2006; pp. 1–7.

33. Berkovsky, S.; Freyne, J. Group-based recipe recommendations: Analysis of data aggregation strategies. In Proceedings of the Fourth ACM Conference on Recommender Systems, Barcelona, Spain, 26–30 September 2010; pp. 111–118.

34. Narducci, F.; Musto, C.; Polignano, M.; de Gemmis, M.; Lops, P.; Semeraro, G. A recommender system for connecting patients to the right doctors in the healthnet social network. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; pp. 81–82.

35. Guo, L.; Jin, B.; Yao, C.; Yang, H.; Huang, D.; Wang, F. Which doctor to trust: A recommender system for identifying the right doctors. *J. Med. Internet Res.* **2016**, *18*, e6015. [CrossRef]

36. Han, Q.; Ji, M.; De Troya, I.M.D.R.; Gaur, M.; Zejnilovic, L. A hybrid recommender system for patient-doctor matchmaking in primary care. In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018; pp. 481–490.

37. Gräßer, F.; Beckert, S.; Küster, D.; Abraham, S.; Malberg, H.; Schmitt, J.; Zaunseder, S. Neighborhood-based Collaborative Filtering for Therapy Decision Support. In Proceedings of the HealthRecSys@ RecSys, Como, Italy, 27–31 August 2017; pp. 22–26.

38. Gmez-Portes, C.; Castro-Schez, J.J.; Albusac, J.; Monekosso, D.N.; Vallejo, D. A fuzzy recommendation system for the automatic personalization of physical rehabilitation exercises in stroke patients. *Mathematics* **2021**, *9*, 1427. [CrossRef]

39. Ishraque, M.T.; Zjalic, N.; Zadeh, P.M.; Kobti, Z.; Olla, P. Artificial intelligence-based cardiac rehabilitation therapy exercise recommendation system. In Proceedings of the 2018 IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA, 5–7 October 2018; pp. 1–5.

40. Ferretto, L.R.; Bellei, E.A.; Biduski, D.; Bin, L.C.P.; Moro, M.M.; Cervi, C.R.; De Marchi, A.C.B. A physical activity recommender system for patients with arterial hypertension. *IEEE Access* **2020**, *8*, 61656–61664. [CrossRef]

41. Zaboleeva-Zotova, A.; Orlova, J.A.; Zubkov, A.V.; Donsckaia, A.R. Intelligent recommendation system for patient rehabilitation. *Artif. Intell. Decis. Mak.* **2024**, *1*, 26–37.

42. Li, J.; Kwong, P.W.; Lua, E.; Chan, M.Y.; Choo, A.; Donnelly, C. Development of a convolutional neural network (CNN) based assessment exercise recommendation system for individuals with chronic stroke: A feasibility study. *Top. Stroke Rehabil.* **2023**, *30*, 786–795. [CrossRef]