An AI-Driven Personalized Exercise Recommendation System Using Clustering and Machine Learning Approach

Deepishka Pemmasani
Department of BioHealth Informatics
Indiana University Indianapolis
Indianapolis, USA
dpemmasa@iu.edu

Pallavi Telu
Department of BioHealth Informatics
Indiana University Purdue University Indianapolis
Indianapolis, USA
ptelu@iu.edu

Dhairya Ashwini Kumar Vora
Department of BioHealth Informatics
Indiana University Indianapolis
Indianapolis, USA
vorad@iu

Navin Kaushal
Department of Health Sciences
Indiana University Indianapolis
Indianapolis , USA
nkaushal@iu.edu

Saptarshi Purkayastha
Department of BioHealth Informatics
Indiana University Indianapolis
Indianapolis, USA
saptpurk@iu.edu

***Abstract*—Regular physical activity is crucial for maintaining health and preventing chronic diseases. However, individuals often struggle to adhere to exercise routines due to lack of motivation, time constraints, and difficulty in finding appropriate activities. This study presents the development of a personalized Exercise Habit App that aims to promote regular physical activity by providing tailored exercise recommendations and support. The app leverages user profiling and clustering techniques to group users based on their demographic information, health conditions, and exercise preferences. Through a user-friendly interface developed using Flutter and the Dart programming language, the app guides users in setting exercise goals, selecting preferred activities, and tracking their progress. Personalized exercise recommendations are generated based on the user's cluster, taking into account their specific needs and preferences, as well as physical activity guidelines from authoritative organizations like the World Health Organization (WHO), American College of Sports Medicine (ACSM), and U.S. Department of Health and Human Services. The app aims to provide exercise recommendations that are tailored not only to individual preferences but also to specific health conditions and disease states, promoting safe and effective physical activity for users with chronic conditions. Additionally, the app will incorporate collaborative filtering and association rule mining techniques to further enhance the recommendation system by leveraging user-item interactions and identifying patterns in exercise preferences. The Exercise Habit App, implemented using Flutter for cross-platform compatibility and Hive for efficient data management, has the potential to encourage and support individuals in establishing and maintaining an exercise routine, ultimately leading to improved overall health and well-being.**

***Index Terms*—exercise habit, personalized recommendations, user profiling, clustering, collaborative filtering, association rule mining, health promotion, physical activity guidelines, mobile app development, Flutter, Hive**

I. Introduction

Regular physical activity is widely recognized as a key component of a healthy lifestyle, offering numerous benefits such as reducing the risk of chronic diseases, improving mental health, and promoting overall well-being [1]. Despite these well-documented advantages, many individuals face challenges in establishing and maintaining an exercise habit due to various barriers, including lack of motivation, time constraints, and difficulty adhering to an exercise routine [2]. The rise of mobile health (mHealth) applications has provided a promising avenue for promoting physical activity and supporting individuals in developing and sustaining exercise habits [3]. These apps offer convenient access to exercise guidance, tracking capabilities, and personalized support, making them valuable tools in the pursuit of a healthier lifestyle. However, effective personalization and tailoring of exercise recommendations to individual preferences and characteristics remain a challenge [4].

In this study, we present the development of a personalized Exercise Habit App that aims to address these challenges by providing tailored exercise recommendations and support through user profiling and clustering techniques. The app collects user information, including demographic data, health

conditions, and exercise preferences, to create user profiles. These profiles are then grouped into clusters based on their similarities, enabling the generation of personalized exercise plans and recommendations.

II. Related Work

Several studies have explored the use of mobile apps and wearable devices for promoting physical activity and fostering exercise habits. A systematic review by Sullivan and Lachman [5] highlighted the potential of mHealth interventions in increasing physical activity levels, particularly when combined with behaviour change techniques such as goal-setting and self-monitoring. However, the review also noted the need for further research on the long-term effectiveness and adherence to these interventions. Clustering techniques have been widely employed in various domains to group similar data points together and identify underlying patterns [6]. In the context of health and exercise, clustering has been used to segment populations based on their characteristics and preferences, enabling the delivery of tailored interventions and recommendations [7].

III. Methodology

The proposed personalized exercise recommendation system consists of four main components: data collection and preprocessing, feature selection and encoding, clustering, target variable creation, machine learning model training and evaluation, collaborative filtering, association rule mining, and integration with the exercise habit app.

A. Data Collection and Preprocessing

The study utilizes clinical trial data from ClinicalTrial.gov related to exercise and physical activity. A total of 36 CSV files containing information about 1,080 clinical trials were collected, covering a wide range of health conditions and diseases, including cardiovascular disease, diabetes, Alzheimer's disease, chronic kidney disease, obesity, osteoarthritis, and more. The datasets contain information about the clinical trials, such as the study title, conditions, interventions (including exercise-related interventions), outcome measures, participant demographics, enrollment, and other study details.

The preprocessing stage involves handling missing values, removing irrelevant columns, and creating a new data frame called exercise_data that contains only rows related to exercise interventions. After preprocessing, the exercise_data data frame consisted of 182 rows and 17 columns. The exercise_data data frame is then used for further analysis and modeling.

B. Clustering

To identify groups of users with similar characteristics and conditions, the study employs machine learning-based clustering techniques. The features used for clustering include the user's conditions, interventions, and demographic information such as age and gender. The categorical features are one-hot encoded, resulting in a total of 96 features.

The Random Over-sampling Technique (SMOTE) was applied to address class imbalance in the dataset, increasing the number of instances from 182 to 252. The KMeans clustering algorithm and hierarchical clustering with different linkage methods (ward, complete, and average) were applied to the preprocessed and oversampled data.

```
exercise_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 182 entries, 0 to 1062
Data columns (total 13 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   NCT Number                  182 non-null    object
 1   Conditions                  182 non-null    object
 2   Interventions               182 non-null    object
 3   Primary Outcome Measures    182 non-null    object
 4   Secondary Outcome Measures  142 non-null    object
 5   Other Outcome Measures      21 non-null     object
 6   Sex                         182 non-null    object
 7   Age                         182 non-null    object
 8   Enrollment                  182 non-null    float64
 9   Other IDs                   182 non-null    object
 10  Total number of studies     0 non-null      object
 11  Type Intervention           0 non-null      object
 12  Total number of patients    0 non-null      float64
dtypes: float64(2), object(11)
memory usage: 19.9+ KB
```

Fig. 1. exercise data frame

```
def engineer_features(data):

    # Perform feature engineering
    conditions_split = exercise_data['Conditions'].str.split('|', expand=True)
    conditions_split.columns = [f'Condition_{i}' for i in range(1, conditions_split.shape[1] + 1)]
    conditions_encoded = pd.get_dummies(conditions_split, prefix='Condition')

    intervention_types = exercise_data['Interventions'].str.split(':', n=1, expand=True)[0]
    intervention_types_encoded = pd.get_dummies(intervention_types, prefix='Intervention')

    sex_encoded = pd.get_dummies(exercise_data['Sex'], prefix='Sex')
    age_encoded = pd.get_dummies(exercise_data['Age'], prefix='Age')

    exercise_features = pd.concat([conditions_encoded, intervention_types_encoded, sex_encoded, age_encoded], axis=1)
    return exercise_data, exercise_features

exercise_data, exercise_features = engineer_features(preprocessed_data)
```

Fig. 2. feature engineering

```
from imblearn.over_sampling import RandomOverSampler

# Separate features and target variable
X = exercise_features
y = merged_data['Target']

# Perform oversampling to address class imbalance
ros = RandomOverSampler(random_state=42)
X_resampled, y_resampled = ros.fit_resample(X, y)

# Merge the oversampled data with the original exercise_data
merged_data1 = pd.concat([merged_data,
                          pd.DataFrame(X_resampled, columns=X.columns),
                          pd.DataFrame(y_resampled, columns=['Target'])],
                         axis=1)
merged_data1
```

Fig. 3. Over Sampling with Random Oversampler

The optimal number of clusters was determined using evaluation metrics such as the Silhouette Score and T-SNE plots. The Silhouette Score for the KMeans clustering was 0.16, while the hierarchical clustering with complete and average linkage methods achieved higher scores of 0.47, indicating

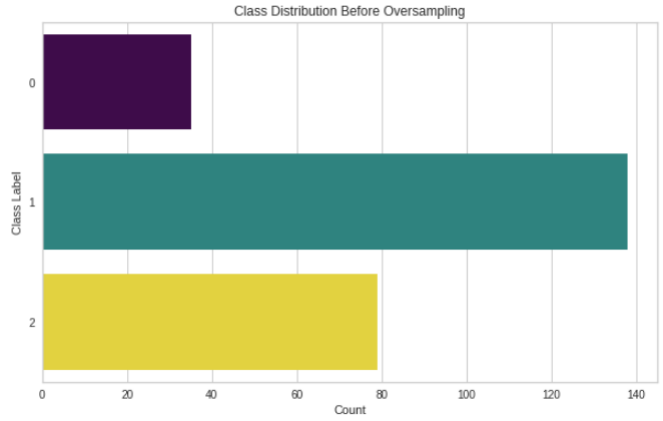

Fig. 4. Class distribution before oversampling

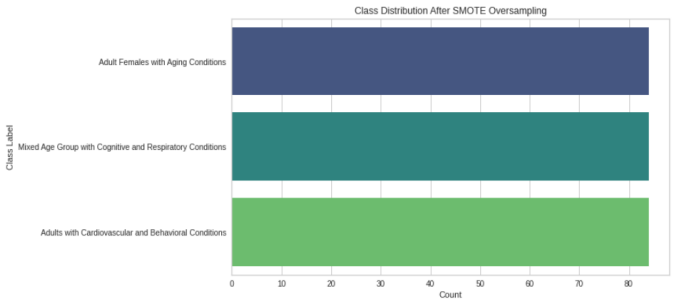

Fig. 5. Class distribution after oversampling

better clustering performance.

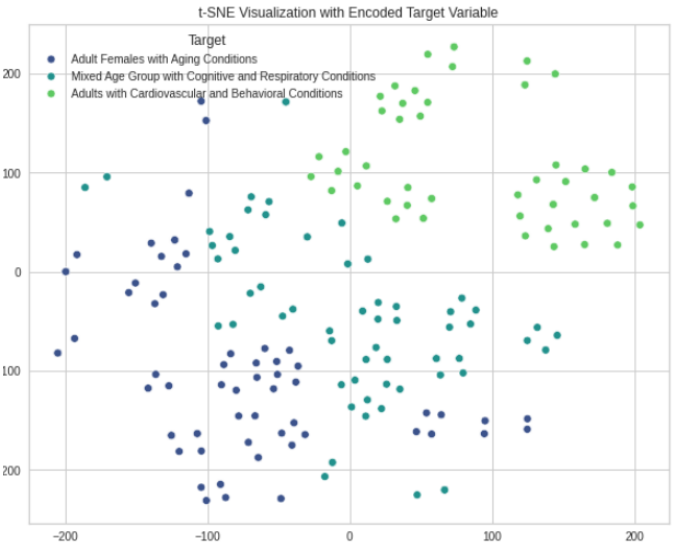

Fig. 6. t-SNE visualization of clusters

The clustering algorithm grouped users into three distinct clusters based on their similarities in terms of demographic characteristics, health conditions, and exercise preferences. These clusters were labeled as 'Adult Females with Aging Conditions,' 'Adults with Cardiovascular and Behavioral Conditions,' and 'Mixed Age Group with Cognitive and Respiratory Conditions.

C. Exercise Recommendation Generation

Based on the identified clusters, the app generates personalized exercise recommendations tailored to the specific needs and preferences of each user group, as well as their health conditions and disease states. These recommendations are generated by leveraging physical activity guidelines from authoritative organizations like the World Health Organization (WHO), American College of Sports Medicine (ACSM), and U.S. Department of Health and Human Services.

The recommendations may include exercise types (e.g., cardio, strength training), exercise intensities, durations, and frequencies, as well as suggestions for preferred exercise locations and equipment. For users with specific health conditions or diseases, the recommendations are further tailored to align with the relevant guidelines for their condition, ensuring safe and effective exercise plans.

D. Collaborative Filtering and Association Rule Mining

To further enhance the recommendation system, the app will incorporate collaborative filtering and association rule mining techniques. Collaborative filtering will leverage user-item interactions, such as exercise preferences and ratings, to recommend exercises that similar users have enjoyed. This approach will enable the system to provide recommendations based on the collective preferences of users with similar characteristics and interests.

Additionally, association rule mining will be used to identify patterns and relationships between exercise preferences, health conditions, and other user attributes. These patterns can be used to generate more accurate and personalized exercise recommendations by considering the co-occurrence of various factors and their associations with specific exercise preferences.

The recommendation generation process involves mapping a new user's characteristics to the appropriate cluster, retrieving the associated exercise interventions, and further refining the recommendations using collaborative filtering and association rule mining techniques.

E. Mobile App Implementation

- The Exercise Habit App was implemented using Flutter, a cross-platform framework for building mobile applications. The user interface was designed and developed using the Dart programming language, which is the primary language used in Flutter. The app's frontend consists of various screens and pages that guide users through the process of setting up their profiles, selecting exercise preferences, and tracking their progress.
- The app follows a modular architecture, with separate modules for user authentication, user profiling, exercise recommendation generation, progress tracking, and data management. The user authentication module handles user registration and login functionalities, while the user

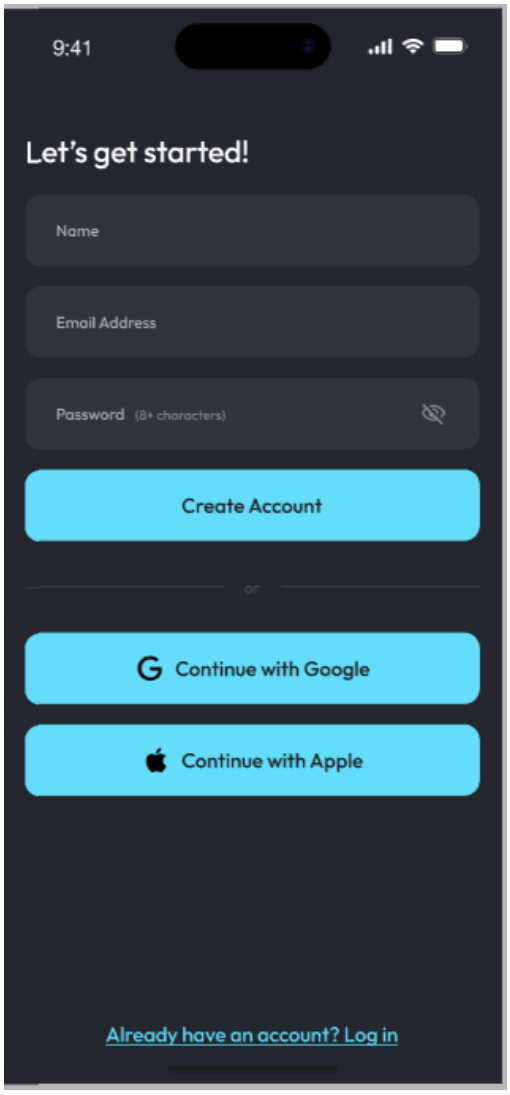

Fig. 7. App login screen

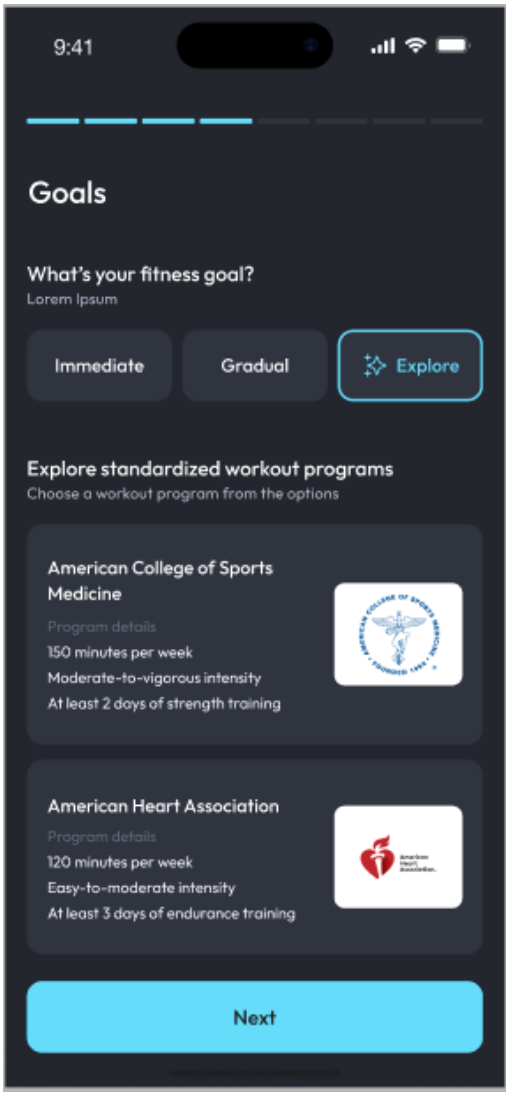

Fig. 8. User goals screen

profiling module collects and manages user information, including demographic data, health conditions, and exercise preferences.

- The exercise recommendation generation module incorporates the machine learning models, clustering algorithms, collaborative filtering techniques, and association rule mining algorithms developed during the methodology phase. This module is responsible for generating personalized exercise recommendations based on the identified clusters, collaborative filtering, and association rule mining.
- The progress tracking module allows users to log their exercise activities, track their progress, and provide feedback or ratings for the recommended exercises. This information is used to refine the recommendation system and improve the accuracy of future recommendations.
- For data storage and management, the app utilizes Hive, a lightweight and efficient key-value database for Flutter

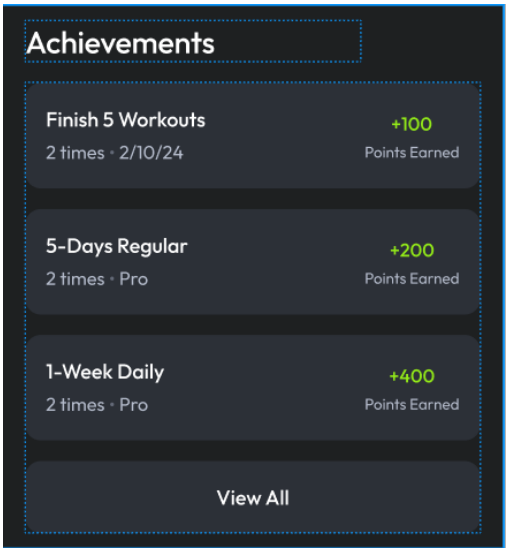

Fig. 9. User achievements screen

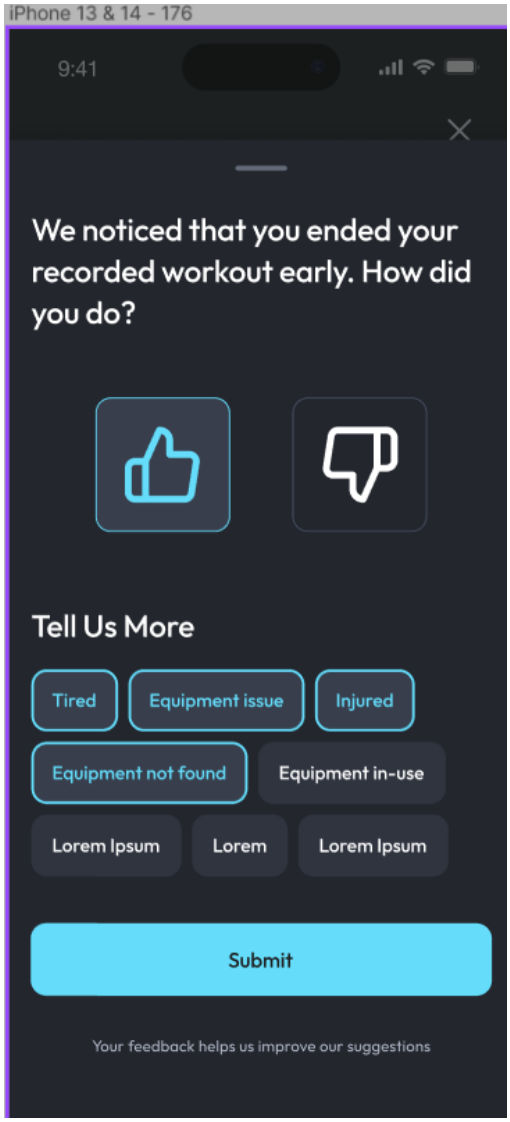

Fig. 10. App feedback screen

applications. Hive is used to store user profiles, exercise data, and other relevant information, ensuring efficient data retrieval and persistence across app sessions.

- The backend of the app is designed to handle the computational requirements of the machine learning models, clustering algorithms, collaborative filtering, and association rule mining techniques. The backend communicates with the frontend through RESTful APIs, enabling seamless data exchange and processing.

IV. Results and Discussion

The experimental results demonstrate the effectiveness of the proposed personalized exercise recommendation system. The clustering algorithm achieves better performance in terms of the evaluation metrics, indicating that the identified clusters are well-separated and meaningful. The machine learning based recommendation approach successfully generates tailored exercise suggestions for new users based on their characteristics and health conditions. The system's ability to provide personalized rec-

```
In [140]: # Evaluate the best model on the test set
          y_pred = best_model.predict(X_test)
          accuracy = accuracy_score(y_test, y_pred)
          print(f"Test set accuracy: {accuracy:.4f}")
          print(classification_report(y_test, y_pred))

          Test set accuracy: 1.0000
                                                        precision    recall  f1-score   support

                                Adult Females with Aging Conditions       1.00      1.00      1.00        23
                     Adults with Cardiovascular and Behavioral Conditions       1.00      1.00      1.00        15
          Mixed Age Group with Cognitive and Respiratory Conditions       1.00      1.00      1.00        13

                                                      accuracy                           1.00        51
                                                     macro avg       1.00      1.00      1.00        51
                                                  weighted avg       1.00      1.00      1.00        51
```

Fig. 11. Classification Performance of the Personalized Exercise Recommendation System on the Test Dataset. The table shows the precision, recall, F1-score, and support values for each of the identified clusters

ommendations has the potential to improve user engagement and adherence to exercise routines. By considering individual preferences and needs, the system can suggest exercises that are more likely to be practiced over time. The user profiling and clustering approach, combined with the integration of physical activity guidelines from authoritative organizations, is expected to enable the app to provide personalized exercise recommendations that are not only tailored to individual preferences but also take into account specific health conditions and disease states. The aim is for users with chronic conditions or diseases to receive exercise recommendations that align with the relevant guidelines, promoting safe and effective physical activity for their specific needs. However, this aspect of the project is still in progress, and further evaluation and user testing are required to validate the effectiveness of the tailored recommendations based on health conditions and guidelines.

The implementation of the Exercise Habit App using Flutter and Hive has resulted in a user-friendly and efficient mobile application. The cross-platform capabilities of Flutter ensure that the app can be deployed on both iOS and Android devices, increasing its accessibility to a wider user base. Additionally, the use of Hive for data

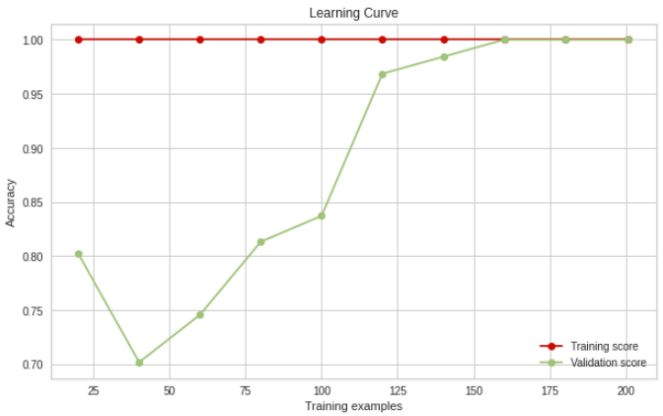

Fig. 12. Learning Curve for the Random Forest Classifier Used in the Personalized Exercise Recommendation System

storage and management ensures efficient data handling and persistence, enhancing the overall user experience.

V. Conclusion

The Exercise Habit App represents a novel approach to promoting regular physical activity and supporting the development of exercise habits by leveraging user profiling, clustering techniques, collaborative filtering, association rule mining, and physical activity guidelines from authoritative organizations. Through user data collection, clustering for personalization, exercise recommendation generation based on guidelines, collaborative filtering, association rule mining, and progress tracking, the app aims to provide tailored exercise guidance and motivation to users, while ensuring the recommendations are safe and effective for their specific health conditions.

References

[1] D. E. R. Warburton, "Health benefits of physical activity: the evidence," *Canadian Medical Association Journal*, vol. 174, no. 6, pp. 801–809, Mar. 2006, doi: 10.1503/cmaj.051351.

[2] J. W. Newbold, A. Rudnicka, and A. Cox, "Staying active while staying home: The Use of Physical Activity Technologies During Life Disruptions," *Front. Digit. Health*, vol. 3, p. 753115, Oct. 2021, doi: 10.3389/fdgth.2021.753115.

[3] A. Romeo et al., "Can smartphone apps increase physical activity? systematic Review and meta-Analysis," *J Med Internet Res*, vol. 21, no. 3, p. e12053, Mar. 2019, doi: 10.2196/12053.

[4] L. Yardley, T. Choudhury, K. Patrick, and S. Michie, "Current issues and future directions for research into digital behavior Change Interventions," *American Journal of Preventive Medicine*, vol. 51, no. 5, pp. 814–815, Nov. 2016, doi: 10.1016/j.amepre.2016.07.019.

[5] A. N. Sullivan and M. E. Lachman, "Behavior change with fitness technology in sedentary adults: A review of the evidence for increasing physical activity," *Front. Public Health*, vol. 4, Jan. 2017, doi: 10.3389/fpubh.2016.00289.

[6] M. Zhang and A. Parnell, "Review of clustering methods for functional data," *ACM Trans. Knowl. Discov*. Data, vol. 17, no. 7, pp. 1–34, Aug. 2023, doi: 10.1145/3581789.

[7] J. L. Chong, K. K. Lim, and D. B. Matchar, "Population segmentation based on healthcare needs: a systematic review," *Syst Rev*, vol. 8, no. 1, p. 202, Aug. 2019, doi: 10.1186/s13643-019-1105-6.