OPEN ACCESS

*International*
*Journal*
*Of Modern Engineering Research (IJMER)*

# Digital Biomarker-Based Clustering of Wearable Data Using Unsupervised Machine Learning for Health Monitoring.

SupriyaSapa- *MCA dept. IMSCDR (Pune University) Ahmednagar , India*
Dr.SharvariTamane – *UDICT dept. MGM University Ch.Sambhajinagar, India,*
ORCidhttps://orcid.org/0009-0008-1631-6970

***Abstract***
*Digital biomarkers, collected through wearable devices such as smart watches and fitness trackers, offer a continuous and non-invasive approach to monitoring human health. This study uses a comprehensive dataset of 556 users tracked over 9 months to 1 year, with data collected from smart watches. Key digital biomarkersincluded are heart rate, blood oxygen saturation (SpO₂), step count, calories burned, exercise time, respiratory rate, and flights climbed; demographic variables such as age and gender are also considered.*
*Unsupervised machine learning techniques, specifically K-means and Gaussian Mixture Model (GMM) clustering are used to categorize the users into four distinct clusters—poor, average,good, and excellent by using the elbow method to find the optimal value of k(4). Cluster assignment was made according to the average values across all digital biomarkers as well asuser age and gender.*
*The performance of both clustering algorithms was evaluated using the silhouette score, which indicated that K-means provided more accurate and well-separated clusters compared to GMM. The study also demonstrated that when new records with average digital biomarker values, age, and gender are given to the K-means model, it assigns to one of the four predefined clusters, enabling automated health status predictions.*
*This research highlights the importance of integrating diverse wearable technologies and key demographic factors with machine learning for advancing biomedical and life sciences research, supporting practical applications in healthseg mentation and dynamic prediction.*
***Keywords****: Digital biomarkers, Wearable devices, Unsupervised Machine Learning, health monitoring.*

---

Date of Submission: 09-04-2026 Date of acceptance: 22-04-2026

---

## I. Introduction

### 1.1 Digital Biomarkers in Biomedical Science

The emergence of wearable devices equipped with embedded sensors has revolutionized health monitoring, enabling real-time tracking of a wide range of physiological and physical parameters, including heart rate, step count, blood oxygen saturation ($SpO_2$), calories burnt, respiratory rate, exercise time, flights climbed, and sleep activity. These measurable parameters, often referred to as digital biomarkers, provide objective insights into an individual's health status. The growing popularity of wearables is driven by their accessibility, ease of use, and ability to provide instant feedback, allowing individuals to take a more active role in managing their health.

In the field of biomedical science, leveraging digital biomarkers is increasingly important for advancing personalized healthcare, disease detection, and preventive interventions. However, a significant challenge lies in effectively harnessing the vast amount of data generated by wearable devices. Accurate interpretation of this data is essential to translate it into actionable health insights. This study explores the application of machine learning techniques to improve the interpretation of digital biomarkers from wearable technology, with a focus on categorizing users into meaningful health profiles.

### 1.2 Wearable Devices

Wearable devices can be broadly classified into categories such as activity trackers, smart watches, medical-grade wearable's, and smart clothing, depending on their primary purpose and functionality. While consumer-grade devices (e.g., Fit bit, Apple Watch, Garmin) focus on general fitness and wellness, clinical-grade wearables are designed for medical monitoring and disease management. Regardless of classification, the reliability of digital biomarkers collected by these devices is highly dependent on correct device usage. For instance, improper placement of a smart watch, loose fitting of fitness bands, or inadequate skin contact can lead to significant deviations in recorded measurements, thereby affecting the accuracy of health assessments.

In the context of biomedical science, wearable devices provide a unique opportunity to continuously monitor digital biomarkers, offering objective and real-time insights into an individual's physiological and health status. However, alongside accuracy, data privacy and security remain major concerns. These devices continuously collect sensitive health information, which, if improperly stored or transmitted, can pose risks of unauthorized access, identity theft, and misuse. Ensuring secure data transmission, encryption, and compliance with privacy regulations is therefore critical in wearable-based biomedical research and personalized healthcare applications.

### 1.3 Machine Learning

Machine learning (ML) is a branch of artificial intelligence that enables systems to learn from data and improve performance without being explicitly programmed. ML algorithms are broadly classified into three categories: supervised learning, unsupervised learning, and reinforcement learning**.**

In the context of health data analysis, unsupervised learning is particularly useful because it can identify hidden structures in datasets where the output variable (ground truth) is not available. One of the most common unsupervised learning techniques is clustering**,** which is used to group unlabelled data into meaningful categories. The main objective of clustering is to organize data points such that those within the same group share similar characteristics, while those in different groups exhibit significant differences.

A crucial aspect of clustering is the choice of a distance metric, which determines how similarity between data points is measured. The most commonly used metric is Euclidean distance, where data points closer together are considered more similar.

Among clustering methods, K-means clustering is widely applied due to its simplicity and effectiveness. In K-means, the dataset is divided into $K$ clusters, with each cluster represented by its centroid. The algorithm iteratively assigns each data point to the nearest centroid, minimizing the Within-Cluster Sum of Squares (WCSS)**,** which measures the squared distance between points and their respective centroids.

The number of clusters ($K$) must be specified in advance, which makes its selection a critical step. To determine the optimal value of $K$, the Elbow Method is often used. This involves plotting the number of clusters against the total WCSS. Initially, WCSS decreases sharply as $K$ increases, but beyond a certain point, the rate of decrease slows, creating a shape resembling an elbow. The point at which this curve bends is considered the optimal number of clusters.

In addition to K-means, the Gaussian Mixture Model (GMM) is another unsupervised learning technique applied in this study. Both K-means and GMM operate on unlabelled data, learning patterns and structures to group data points into clusters. While K-means assumes spherical clusters of equal size, GMM provides greater flexibility by allowing clusters of different shapes and distributions. These methods are particularly well-suited for analysing wearable-derived digital biomarkers, where the underlying health profiles may not be predefined but can be revealed through clustering.

## II. Problem Statement

Wearable devices such as smart watches and fitness bands generate vast amounts of physiological and activity-related data, including heart rate, step count, respiratory rate, exercise time, flights climbed calories burnt, and blood oxygen saturation. While these digital biomarkers have significant potential for health monitoring, the raw data are often noisy, inconsistent, and difficult for users to interpret meaningfully. Many existing studies analyse wearable data in isolation or with limited machine learning methods, which restricts the accuracy and reliability of health predictions. Moreover, challenges such as device placement accuracy, variations across brands, and concerns regarding data privacy and security further complicate the effective use of wearable technology in healthcare. There is a clear need for a robust, reliable, and interpretable approach that can integrate multiple features from wearable devices to predict an individual's health status more accurately.

## III. Literature Review

- Walk, talk, think, see and feel: harnessing the power of digital biomarkers in healthcare
- Author: Dylan Powell(2024):
- Journal: npj Digital Medicine
  The digital biomarkers across domains - physical, speech, cognitive, visual, and emotional—have the potential to revolutionize healthcare by enabling continuous, real-world health monitoring. Introducing a framework built on walk, talk, think, see, feel constructs, the author advocates for integrating multiple signals into a single digital biomarker fingerprint to better detect and manage complex health states. Powell also calls attention to key challengesespecially around data representativeness, privacy, and aligning innovations with healthcare's broader goalsunderscoring the need for a balanced, inclusive, and value-driven approach.
- Verification, Analytical Validation, and Clinical Validation (V3): The Foundation of Determining Fit-for-Purpose for Biometric Monitoring Technologies (BioMeTs)
  Authors : Goldsack, J., Coravos, A., Bakker, J., Bent, B., Dowling, A., Fitzer-Attas, C., Godfrey, A., Godino, J., Gujar, N., Ismailova, E., Manta, C., Peterson, B., Vandendriessche, B., Wood, W., & Wang, K. (2020).
  Journal:npj Digital Medicine
  The V3 framework—Verification, Analytical Validation, and Clinical Validation—provides a structured approach for evaluating Biometric Monitoring Technologies (BioMeTs) and their digital measures. Proposed by the Digital Medicine Society, the framework addresses the need for standardization in the rapidly evolving field of digital health. Verification ensures sensor hardware accurately captures raw signals, analytical validation evaluates whether algorithms correctly process those signals into reliable measures, and clinical validation confirms that the resulting digital endpoints are meaningful in real-world patient populations. By creating a common language and evidence pathway, the V3 framework enhances trust, reproducibility, and regulatory acceptance of BioMeTs, making it an essential reference for any discussion of digital biomarker validation
- Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review
  Authors: Grace Shin, Mohammad Hossein Jarrahi, Fei Y, Amir Karami, Nicci Gafinowitz, Ahjung Byun,
  Journal:JournalofBiomedicalInformatics
  Volume & Article ID:Volume 93, Article 103153 (May2019)

Wearable activity trackers (WAT) are electronic monitoring devices that enable users to track and monitor their health related physical fitness metrics including steps taken, level of activity, walking distance, heartrate, and sleep patterns. Objectiveof the study is to understand the rich human-information interaction that is enabled by WAT adoption. In the study Topic modeling methods were used to identify six key themes of WAT research, namely Technology Focus, Patient Treatment and Medical Settings, Behavior Change, Acceptance and Adoption, Selfmonitoring Data Centered, and Privacy.This work raises interdisciplinary awareness about the current landscape of WAT use and the related diversity of interesting research opportunitiesand challenges. The study suggests that WAT devices are multi-dimensional technologies with complex impacts. Understanding WAT and their technologicaland non-technological aspects requires various research perspectives.This multi-dimensional framework highlights that WATs are not merely fitness devices but are embedded in complex human-information interactions. The review emphasizes the need for interdisciplinary approaches to fully understand the technological, behavioral, and ethical dimensions of WATs, and opens avenues for future research in both health informatics and user-centered design.

- Wearable Sensors for Remote Health Monitoring

Authors**:** SumitMajumder, Tapas Mondal and M. Jamal Deen - Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada;
Journal**:**Sensors (an open-access journal published by MDPI)-2017
Remote health monitoring, based on non-invasive and wearable sensors, modern communication and information technologies offers an efficient and cost-effective solution that allows the elderly to continue to live in their comfortable home environment instead of expensive healthcare facilities. These systems will also allow healthcare personnel to monitor important physiological signs of their patients in real time, assess health conditions and provide feedback from distant facilities. In this paper, they have presented and compared several

low-cost and non-invasive health and activity monitoring systems. Finally, compatibility of several communication technologies as well as future perspectives and research challenges in remote monitoring systems are discussed.

- Machine Learning for Healthcare Wearable Devices:The Big Picture

Authors**:** Farida Sabry, Tamer Eltaras, Wadha Labda, Khawla Alzoubi, Qutaibah Malluhi
Journal:Journal of Healthcare Engineering (Hindawi)-2022
The paper highlightsthe Machine Learning Techniques used, the different modalities used, and theavailable data sets and the different challenges facing machine learning applications on wearable devices like deployment alter natives, power consumption, storage and memory, utility and user acceptance, data availability and reliability, c ommunication, security and privacy were discussed while identifying possible solutions found in the literature. The objective of the study is to highlight the various ML techniques and the challenges in the deployment of wearable devices. The methods used are applied on datasets available for human activity recognition. K-NN,SVM,LR,Tree-based,Deep learning models are used for analysis of the data. Further research concerning data availability, reliability, and privacy to enable effective and efficientlearning from data generated by wearable devices. The w earable devices are used for remote patient monitoring and detection of any irregularities with the human body

## IV. Research Objectives

The main objectives of this research are:

1. To integrate digital biomarkers with demographic factors such as age and gender to enable comprehensive health data analysis.
2. To apply unsupervised machine learning techniques to identify natural groupings within health-related digital biomarker data.
3. To develop a predictive framework for classifying new digital biomarker profiles into predefined clusters, supporting personalized health monitoring.

## V. Research Methodology

This research adopts anunsupervised machine learning-based clustering methodology to evaluate and classify the users into a health category by forecasting an individual's yearly health score based on physiological and activity data derived from wearable devices. In this research, wearable-derived features such as heart rate, step count, $SpO_2$, Respiratory rate, flights climbed and calories burnt are employed to predict a health category. Since this category reflects the current health state (diagnostic role) and potential future health risks (prognostic role), the digital biomarkers used in this work can be classified under both diagnostic and prognostic categories. Data Analysis is dine using Python language.

### 5.1 Data Collection Methodology

In this study, wearable device data was collected from 556 individuals over a period ranging from 9 to 12 months. For each participant, a separate file was maintained containing daily health records**.** To ensure consistency and reduce variability, the data was aggregated on a monthly basis.This aggregation was performed by applying a group-by operation on the month and calculating the average values for each health parameter (e.g., step count, heart rate, $SpO_2$, calories burnt, respiratory rate, flights climbed, etc.). As a result, each participant's dataset was transformed into a monthly summary comprising mean values for all parameters.Following this transformation, a single-row summary per user was generated to represent their overall health trends. The individual records of all 556 participants were then combined into a single dataset. This structured dataset served as the foundation for applying machine learning techniques for clustering the users into different heath categories based on digital biomarkers and prediction for a new user data.

### 5.2 Data from Wearable Devices

Wearable devices from different brands provide a diverse range of health-related parameters, varying in both type and quantity. However, certain parameters are commonly available across most brands and are considered as the key digital biomarkersfor effective health tracking. These include:

- Step Count (steps)
- Walking + Running Distance (km)
- Walking Speed (km/hr)
- Walking Step Length (cm)
- Active Time (min)
- Exercise Time (min)
- Flights Climbed (count)

- Blood Oxygen Saturation (%)
- Respiratory Rate (count/min)
- Heart Rate [Min] (bpm)
- Heart Rate [Max] (bpm)
- Heart Rate [Avg] (bpm)
- Resting Heart Rate (bpm)
- Sleep Activity

These parameters form the core dataset in this study, as they represent reliable indicators of daily activity, cardiovascular health, respiratory efficiency, and overall well-being.

**5.3 Feature Engineering**

In consultation with healthcare professionals and based on medical recommendations, the following parameters were selected as the most relevant features for analysis. These parameters were chosen for their strong association with overall health status and their reliability as digital biomarkers:

- Gender
- Age
- Step Count
- Calories Burnt
- Blood Oxygen Saturation ($SpO_2$)
- Average Heart Rate
- Respiratory Rate
- Exercise Time
- Flights Climbed

To further justify the selection of these features, reference was made to the World Health Organization (WHO) Global Strategy on Physical Activity and Health (2004), which provides internationally recognized health standards to promote physical, mental, and social well-being. For adults aged 18 to 65 years, WHO suggests the following benchmarks:

- Step Count: A common guideline is to aim for 10,000 steps per day.
- Heart Rate: A normal resting heart rate typically falls between 60 and 100 beats per minute.
- $SpO_2$ (Blood Oxygen Saturation): A healthy $SpO_2$ level is generally above 95%.
- Calories Burnt: Adults should aim to burn approximately 300–500 calories per day through moderate physical activity.

By aligning the selected features with WHO health standards and clinical recommendations, this study ensures that the analysis is based on globally accepted health benchmarks, thereby enhancing the validity of the results.

**5.4 Exploratory Data Analysis (EDA)**

Exploratory Data Analysis (EDA) was conducted using Microsoft Power BI to examine the distribution, trends, and interrelationships among the collected health parameters. Along with physiological data obtained from wearable devices, demographic variables such as age and gender were also incorporated into the analysis.

A variety of interactive visualizations—including bar charts, combo charts, dot plots, and box plots—were employed to:

- Explore patterns across different age groups and gender categories,
- Identify correlations between digital biomarkers and demographic factors, and
- Examine their association with overall health outcomes.

This step provided valuable insights into the dataset, enabling a deeper understanding of variability across participants and guiding the selection of appropriate machine learning models for further analysis.

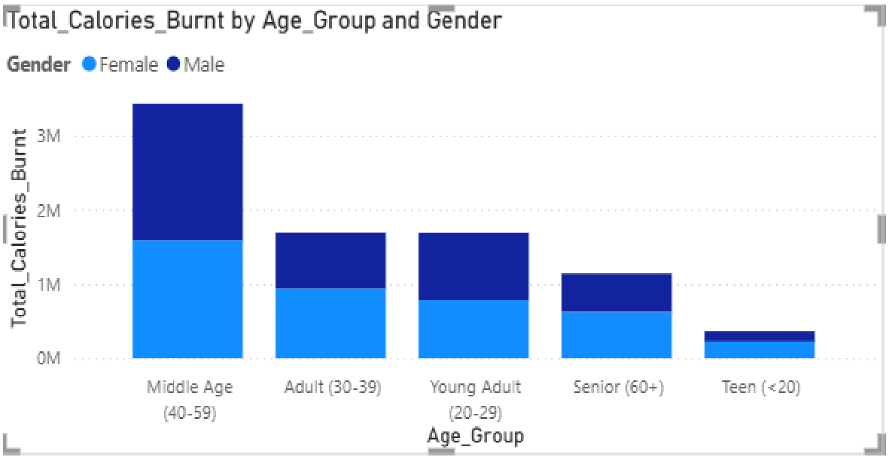


**Fig 1 Stacked bar chart- Total Calories Burnt by Age and Gender**

Total calories burnt are higher in age group 40 to 60 with males having higher then females.

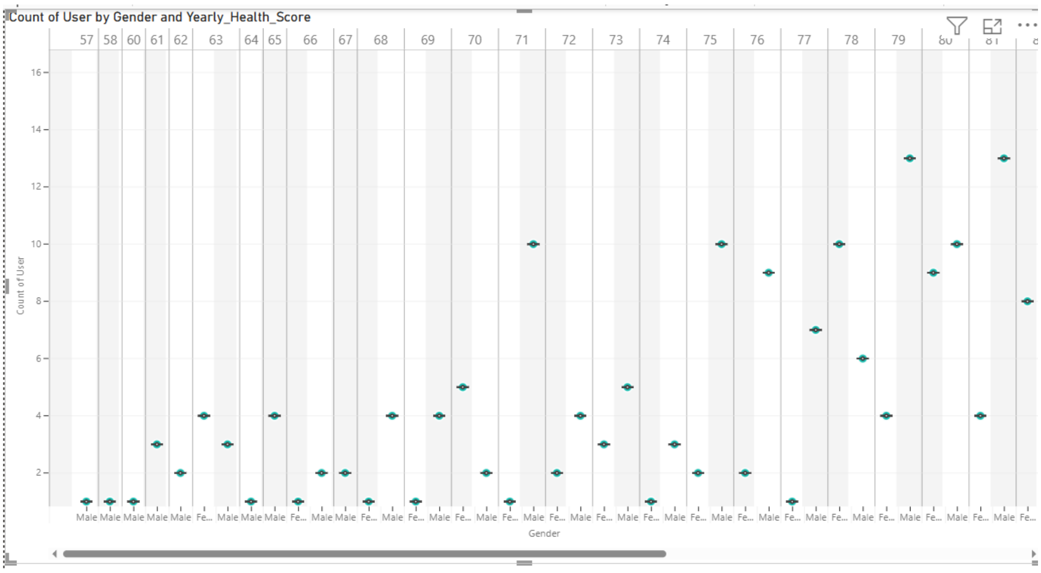


**Fig 2 – Dot Plot - Count of User by Gender and Health Score**

Count of users is higherin males with health score ranging between 75 to 82.

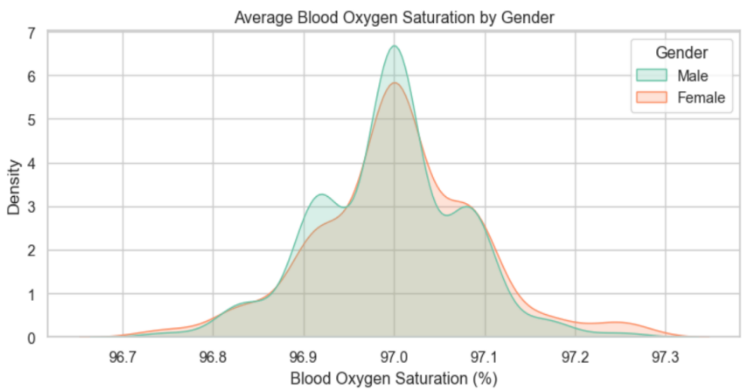


**Fig 3 - Kernel Density Estimate (KDE) - Average Blood Oxygen Saturation by Gender**

To explore gender-wise patterns**,** Kernel Density Estimate (KDE) plot (also called a density plot) were created for Average Blood Oxygen Saturation (%) and Average Heart Rate. The visual highlighted that males had a slightly wider range and higher blood oxygen saturation compared to females.

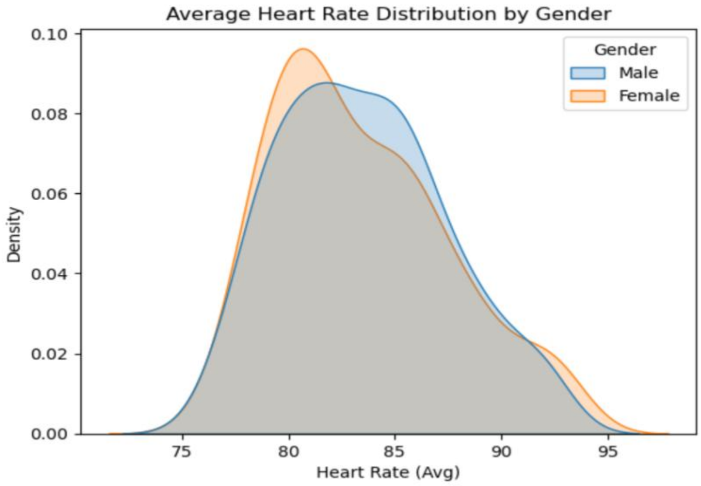


**Fig 4 - Kernel Density Estimate (KDE) - Average Heart Rate Distribution by Gender**

The visual highlighted that heart rate distributions in females is slightly higher than males keeping similar across both groups.

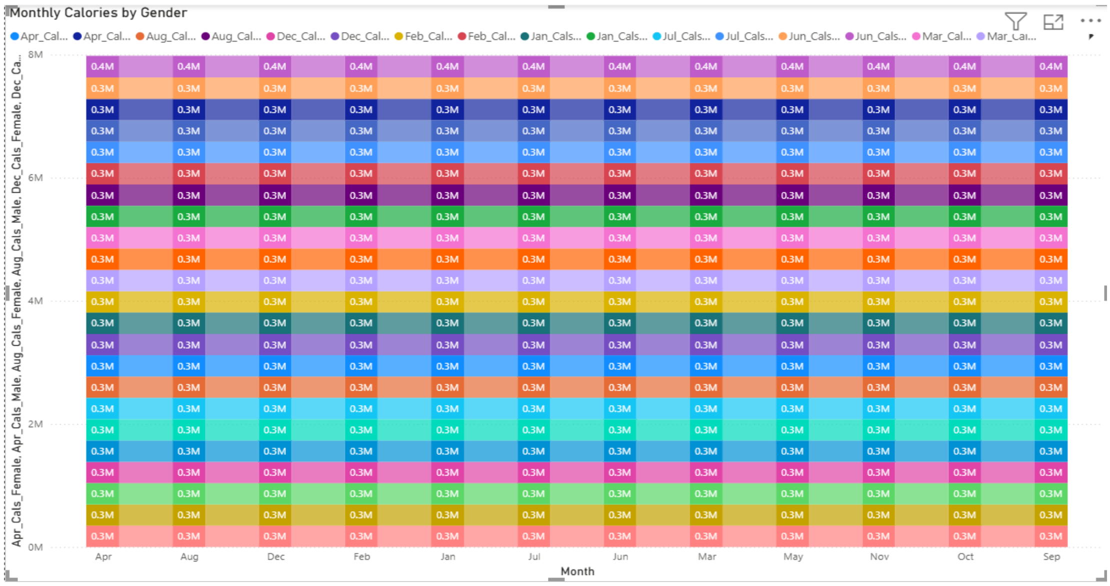


**Fig 5 – Heatmap Style Matrix - Monthly Calories Burnt by Gender**

A correlation heatmap examine interdependencies among variables. It was observed that Avg_Heart_Rate and Total_Calories_Burnt showed moderate positive correlation with the health category, indicating their significance in predicting overall wellness.

## VI. Data Analysis

To analyse the collected dataset, the Elbow Method was first applied to determine the optimal number of clusters ($k$). By plotting the Within-Cluster Sum of Squares (WCSS) against different values of $k$, the curve exhibited a clear inflection point (elbow) at $k = 4$**,** indicating that four clusters provided the best balance between compactness and separation.

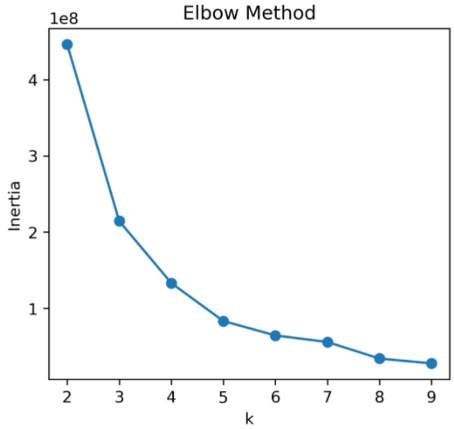


**Fig 6- Optimal number of clusters (elbow point): 4**

### 6.1 K-means Clustering

K-means clustering algorithm was applied with $k = 4$. The algorithm partitioned the dataset into four distinct clusters, with each cluster characterized by a unique cluster centroid representing the average values of the selected features (step count, calories burnt, heart rate, $SpO_2$, Respiratory rate, flightsclimbed, age, and gender). These centroids provide interpretable summaries of the underlying health profiles within the population.

**Cluster Labels Mapping**

{0: 'Poor', 2: 'Average', 3: 'Good', 1: 'Excellent'}

This clustering step enabled the identification of natural groupings in the dataset, thereby facilitating the categorization of individuals into distinct health categories.

**Clustering Results and Health Categorization**

Based on the K-means clustering with $k = 4$, each user in the dataset was assigned to one of the four clusters. To enhance interpretability, the clusters were mapped to health categories(Poor, Average, Good, Excellent) according to the cluster centroids. This mapping illustrates how cluster analysis helps translate raw digital biomarker data into actionable health categories**,** allowing for a more intuitive understanding of individual and population health status.

A sample of the user data with their Cluster ID and corresponding Health Category is shown below:

| | User | cluster_id | health_label |
|---|---|---|---|
| 0 | user1 | 2 | Average |
| 1 | user2 | 0 | Poor |
| 2 | user3 | 0 | Poor |
| 3 | user4 | 2 | Average |
| 4 | user5 | 2 | Average |
| .. | ... | ... | |
| 551 | user552 | 1 | Excellent |
| 552 | user553 | 3 | Good |
| 553 | user554 | 3 | Good |
| 554 | user555 | 3 | Good |
| 555 | user556 | 3 | Good |

**PCA-Based Cluster Visualization**

To visualize the clustering results from the high-dimensional dataset, Principal Component Analysis (PCA) was applied to reduce the feature space to two principal components. This dimensionality reduction preserved the maximum possible variance in the data while enabling a 2D visualization of the four clusters. The resulting PCA scatter plot clearly shows the separation of data points into distinct clusters, each corresponding to a specific health category (Poor, Average, Good, and Excellent). Cluster centroids were also plotted to represent the mean position of each group in the reduced feature space. This visualization demonstrates that the clusters formed by K-means were not random but instead reflect meaningful groupings of participants based on their digital biomarkers and demographic features (age, gender)**.**

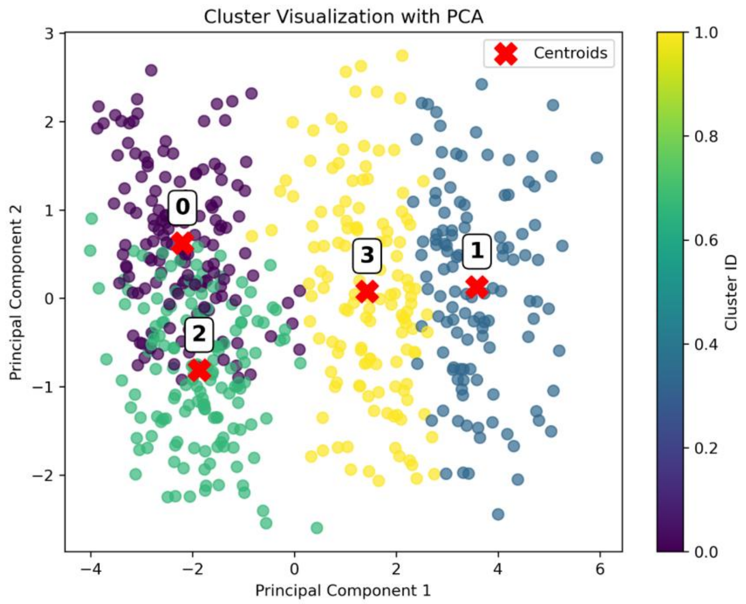


**Fig 7 - PCA-based cluster visualization of a high-dimensional dataset, reduced to 2D.**

**6.2 Gaussian Mixture Model (GMM)-Based Cluster Analysis**

To further explore the underlying structure of the dataset, Gaussian Mixture Model (GMM) clustering was applied. Similar to the previous analysis, the dataset was reduced to two dimensions using PCA for visualization. The GMM clustering results show four distinct clusters, corresponding to thesame health categories (Poor, Average, Good, Excellent)**.** Each data point is associated with a probability of belonging to one cluster.

**GMM Cluster Centres and Visualization**

After applying Gaussian Mixture Model (GMM) clustering to the dataset, four clusters were identified. Each cluster is represented by a cluster centre (mean of the Gaussian component), summarizing the central tendency of the features (step count, calories burnt, heart rate, $SpO_2$,flights climbed, respiratory rate, exercise time age, gender) for that group.

The cluster centres provide interpretable summaries of the underlying health profiles and were used to map each cluster to a health category (Poor, Average, Good, Excellent)**.**

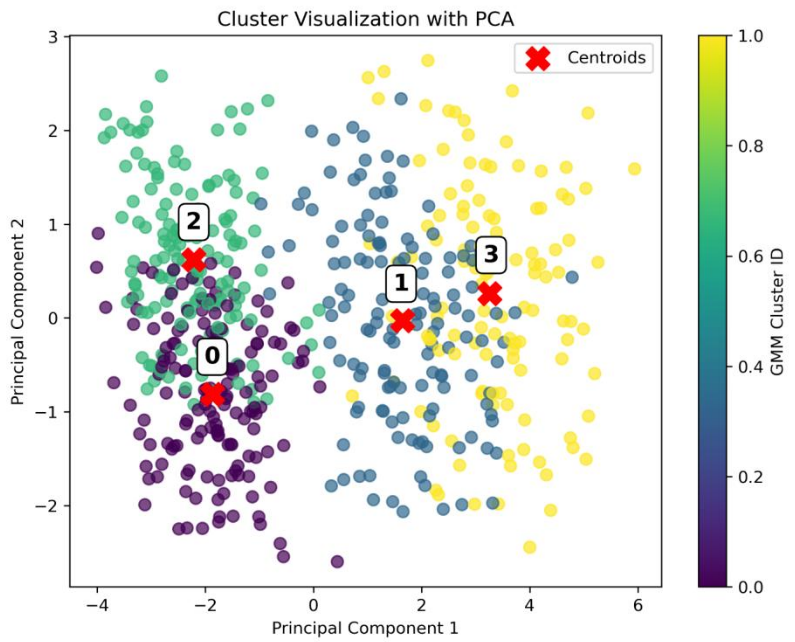


**Fig 8 - PCA-based cluster visualization reduced to 2D for GMM.**

**6.3 Comparison-K-means performs better than GMM in terms of cluster cohesion and separation**

| Model | Silhouette Score |
|---|---|
| K-means | 0.245 |
| GaussianMixture | 0.209 |

## VII. Results and Outcome

- K-means Silhouette Score = 0.245
- GMM Silhouette Score = 0.209

Since 0.245 > 0.209, K-means clustering is performing better in terms of cluster cohesion and separation on your data. Although Gaussian Mixture Model provides flexibility in modelling elliptical clusters, the Silhouette Score indicates that K-means (0.245) produced slightly more cohesive and well-separated clusters compared to GMM (0.209) for the given wearable health dataset. This suggests that the data distribution aligns better with the assumptions of K-means.

The study also demonstrated that when new records with average digital biomarker values, age, and gender are given to the K-means model, it assigns to one of the four predefined clusters, enabling automated health status predictions. In the example given below the new data is predicted to belong to cluster 1.

new data = np.array([[ 1.00000000e+00, 6.50448634e-01, 9.02242660e-01, 5.92100637e-01,1.15477753e-02, -1.00527649e+00,2.636909,12.462862 , 17.699516, 7.58014770e-01]])
# Predict clusters
clusters = kmeans.predict(new data)
print(clusters) # Example output: [1]

## VIII. Conclusion

This study highlights the utility of clustering techniques in analyzingdigital biomarkers derived from wearable devices such as step count, heart rate, $SpO_2$, and calories burned. By applying K-means and Gaussian Mixture Models (GMM), we identified distinct health-related clusters within the user population. Comparative evaluation using silhouette scores indicated that K-means achieved better cohesion and separation (0.245) compared to GMM (0.209), suggesting it is more effective in uncovering meaningful subgroup patterns in this dataset.

The clustering outcomes demonstrate that unsupervised learning can segment individuals into groups that reflect different health profiles, thereby supporting the potential use of digital biomarkers for population-level health monitoring and personalized interventions. These findings reinforce the importance of clustering as an exploratory tool in biomedical science, enabling the transformation of raw wearable data into actionable insights. Future work should expand the feature set (e.g., sleep, stress, mood indicators) and explore multimodal clustering approaches to enhance the interpretability and clinical relevance of digital biomarker research.

## References

[1]. Powell, D. (2024).Walk, talk, think, see and feel: harnessing the power of digital biomarkers in healthcare. npj Digital Medicine, 7, Article 45. https://doi.org/10.1038/s41746-024-01023-w Nature

[2]. Goldsack, J., Coravos, A., Bakker, J., Bent, B., Dowling, A., Fitzer-Attas, C., Godfrey, A., Godino, J., Gujar, N., Ismailova, E., Manta, C., Peterson, B., Vandendriessche, B., Wood, W., & Wang, K. (2020).Verification, Analytical Validation, and Clinical Validation (V3): The Foundation of Determining Fit-for-Purpose for Biometric Monitoring Technologies (BioMeTs). npj Digital Medicine, 3, Article 55. https://doi.org/10.1038/s41746-020-0260-4

[3]. Salman, H. A., Kalakech, A., &Steiti, A. (2024).Random Forest Algorithm Overview.*Babylonian Journal of Machine Learning*, 2024, 69–79. https://doi.org/10.58496/BJML/2024/007

[4]. Ali, J., Khan, R., Ahmad, N., &Maqsood, I. (2012).Random Forests and Decision Trees.*ResearchGate*.https://www.researchgate.net/publication/259235118

[5]. Chen, Y., Wang, J., Yu, C., Gao, W., & Qin, X. (2021). *FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare* (arXiv:1907.09173). arXiv. https://doi.org/10.48550/arXiv.1907.09173

[6]. Chiang, P.-H., Wong, M., & Dey, S. (2021). Using Wearables and Machine Learning to Enable Personalized Lifestyle Recommendations to Improve Blood Pressure. *IEEE Journal of Translational Engineering in Health and Medicine*, *9*, 1–13. https://doi.org/10.1109/JTEHM.2021.3098173

[7]. Neshitov, A., Tyapochkin, K., Kovaleva, M., Dreneva, A., Surkova, E., Smorodnikova, E., & Pravdin, P. (2023). Estimation of cardiorespiratory fitness using heart rate and step count data. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-43024-x

[8]. Odhiambo, C., Wright, P., Corbett, C., & Valafar, H. (n.d.). *MedSensor: Medication Adherence Monitoring Using Neural Networks on Smartwatch Accelerometer Sensor Data*.

[9]. Prasad Raju, N. V. D. S. S. V., & Devi, P. N. (2024). A Comparative Analysis of Machine Learning Algorithms for Big Data Applications in Predictive Analytics. *International Journal of Scientific Research and Management (IJSRM)*, *12*(10), 1608–1630. https://doi.org/10.18535/ijsrm/v12i10.ec09

[10]. Sabry, F., Eltaras, T., Labda, W., Alzoubi, K., & Malluhi, Q. (2022). Machine Learning for Healthcare Wearable Devices: The Big Picture. *Journal of Healthcare Engineering*, *2022*, 1–25. https://doi.org/10.1155/2022/4653923

[11]. Yusuf, A., Jaber, T. A., & Gordon, N. (2025). Comprehensive Health Tracking Through Machine Learning and Wearable Technology. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewjdsis52023588