**Health Prediction Based on Day-to-Day Life Activity using Machine Learning Approach**

Razeena Begam1   
Computer Science (AI&ML) Department,Lakireddy Bali Reddy College of Engineering, Mylavaram, India  
[guide@gmail.com](mailto:guide@gmail.com)

Bhimavarapu Harshini4  
Computer Science (AI&ML) Department,Lakireddy Bali Reddy College of Engineering, Mylavaram, India  
<harshubhimavarapu24@gmail.com>  
Jonnadula Malleswari2  
Computer Science (AI&ML) Department,Lakireddy Bali Reddy College of Engineering, Mylavaram, India  
<monumalleswari@gmail.com>

Tummala Jaya Vardhan5  
Computer Science (AI&ML) Department,Lakireddy Bali Reddy College of Engineering, Mylavaram, India  
[jayavardhan196@gmail.com](mailto:jayavardhan196@gmail.com%20) Vasireddy Aarya Chowdary3  
Computer Science (AI&ML) Department,Lakireddy Bali Reddy College of Engineering, Mylavaram, India  
<aaryavasireddy58@gmail.com>

*Abstract*— The potential of health surveillance systems to offer thorough insights into a person's general well-being is now limited by their limits especially those that only analyze certain illnesses like Parkinson's disease. These methods ignore important health indicators that are necessary for a comprehensive assessment of an individual's health, such as blood pressure, stress levels and heart health. This work is to create a complex framework that analyzes a broad variety of health data gathered from various smartwatch users in order to solve this issue and provide a thorough understanding of several health aspects. The system combines these many health parameters using state-of-the-art data analysis and representation technologies to provide tailored suggestions based on past trends assisting people in making well-informed decisions for proactively managing their health. K-Means and Agglomerative Clustering methods are used to sort people into groups based on health characteristics after the data has been processed using the Standard Scaler and separated into testing and training sets. The ideal number of groups was found using the Elbow Method and K-Means outperformed Agglomerative Clustering in terms of silhouette score with 0.85. The system's user-centric design facilitates the determination of areas for development and promotes a culture of preventative healthcare by offering extensive insights into health patterns. By encouraging proactively rather than reaction healthcare measures this initiative ultimately seeks to empower users with practical information improving both individual health management and social well-being.

Keywords— Agglomerative Clustering, Data Visualization, Health Monitoring, K-Means Clustering, Preventive Healthcare, Smartwatch Data, Standard-Scaler and Wearable Technology.

# Introduction

Recent developments in mobile devices and the increased focus on individual well-being have led to a substantial evolution in health monitoring systems. The scope and usefulness of many systems are limited by their exclusive emphasis on particular ailments, like Parkinson's disease even if there are gadgets like smartwatches [1] that can gather useful health data. The current approach is unable to provide an entire picture of an individual's general health since it ignores important signs like blood pressure, heart rate and stress levels. This lack of broad health monitoring limits possibilities for proactive health management by preventing users from gaining deep insights into their well-being. Therefore, there is an urgent need for a system that can assess several health variables and offer timely helpful advice that encourages preventive measures [2].

Health statistics demonstrate the increasing significance of ongoing health monitoring in the fast-paced world of today. Pursuant to the World Health Organization (WHO) heart disease [3] continues to be the top cause of mortality worldwide, accounting for around 18 million deaths each year. Other diseases such as diabetes, high blood pressure and stress-related issues are becoming more common and affect millions of people every day. For example tension is now an important risk factor to a number of diseases including high blood pressure and heart disease. There is a remarkable chance for technology that is wearable like smartwatches to measure and manage these conditions in real time. By using the huge quantity of data that these devices gather to find warning signs in advance people may take preventive measures before minor health issues become serious ones. Moreover, hypertension is a major cause of early mortality worldwide affecting an estimated 1.28 billion individuals between the ages of 30 and 79.

Nearly 40% of individuals have excessive blood pressure frequently without being aware of it based to the latest WHO research [4]. In a similar vein millions of people suffer from anxiety, sleeplessness and depression, all of which are linked to stress-related health problems. These figures highlight how crucial it is to have a system that can monitor and analyze health data in real time enabling people to better understand their conditions and take prompt action. Through continuous monitoring of vital signs such as blood pressure, heart rate and stress levels wearable technology offers a chance for better health management [5] by revealing important health patterns. The situation is just as worrying in India. The Worldwide Cost of Disease Study states that almost 28% of all deaths in the nation are attributable to cardiovascular illnesses making them the top cause of death. About 25% of Indian people in rural areas and 33% of those in urban areas [6] suffer from hypertension and stress-related disorders are on the rise due to fast-paced lives and increasing urbanization. The demand for people to take charge of their health is growing and wearable health technology is a crucial tool in a healthcare system that is frequently overburdened by the population.

Smartwatches are the most widespread category in the fast grow in Indian wearable industry [7]. This creates new opportunities for the creation of cutting-edge health surveillance systems that can address the unique medical requirements of Indian consumers. Given these alarming statistics there is strong reason that supports the development of a more thorough health monitoring system. This study aims to close the gaps in the current approaches to health monitoring by looking at a wider range of health metrics and avoiding a specialized focus on Parkinson's disease. Through the use of smartwatch data collected from a wide user base this work seeks to create a comprehensive system that can evaluate high blood pressure, levels of stress, heart health and other critical health factors. The technology [8] will not only evaluate the data but also provide users with useful insights that encourage a preventive approach to healthcare. This might significantly reduce the burden on healthcare systems in countries such as India, where preventive care is often ignored.

# literature survey

The development and use of personal health technology has significantly increased over the last ten years, mostly for the purpose of monitoring certain illnesses or medical problems. In order to give users a broad picture of their daily exercise and rest early systems concentrated on monitoring fundamental parameters like blood pressure, step measure and sleep patterns. These technologies however frequently lacked the capacity to detect possible hazards or provide comprehensive insights [9] into general health. Mohammad et al. in today's hectic environment health monitoring is essential since many people develop health problems early in life as a result of their everyday activities [10]. Current approaches use electronic health records to forecast illnesses. A suggested approach predicts a person's general health state including their level of physical activity, food consumption and sleep using machine learning. Additionally, the system tracks health and notifies users when something is out of the ordinary. Thakur et al. using wrist-worn IMU sensors, this study attempts to create a machine learning-based modelling framework for precisely detecting smoking activity among daily tasks. The inexpensive gadget [11] gathers unprocessed sensor data extracts characteristics and creates models for multi-class categorization when forecasting smoking behavior the models had a 98.7% predictive accuracy.

More specialized health monitoring systems emerged as wearable technology advanced with some systems concentrating on particular illnesses like diabetes Parkinson's disease or cardiovascular disorders. Numerous initiatives that focused on employing wearable technology to measure movement, identify tremors and evaluate the disease's development arose. Similar to these systems created for the monitoring of cardiac illness started incorporating electrocardiogram (ECG) capabilities [12] into wearable technology, but their reach was still constrained frequently ignoring the necessity of blood pressure or stress monitoring. Meng et al. in this study 182 patients with stable ischemic cardiovascular disease (SIHD) had their activity tracker data utilized to identify patient-reported outcomes (PROs) using machine learning algorithms [13]. A concealed Markov model and a classifier based on random forests were used to construct the models. The findings shown that health state may be categorized over time using activity tracker data allowing for early intervention. According to the study, activity trackers may be used to follow patient results in real time highlighting the technology's potential for health monitoring.

More extensive health monitoring systems are now possible because to developments in data analysis and machine learning in recent years. Numerous studies have tried to combine various health indicators such blood pressure, stress levels and heart rate variability to give a more comprehensive picture of a person's wellbeing. This systems ability to offer specific and useful information was limited by their frequent use of basic statistical techniques or clustering algorithms [14]. Many of these prior initiatives did not fully explore the potential of real-time monitoring and predictive analytics which limited their effectiveness in encouraging preventive healthcare. Due to limitations on the number of clusters or health metrics included those systems that first used clustering algorithms typically did not provide a thorough dimensional evaluation of user health. This highlights the need for more complex multidimensional approaches that can satisfy the various health needs of modern customers.

# DATA COLLECTION & DATA PREPROCESSING

A wide range of health measures such as blood pressure, heart rate, stress levels and aerobic activity were recorded by the smartwatches that were used to gather the data for this research. Throughout the day these smartwatches continually monitored the user vital signs giving them up-to-date health status information. To ensure that the data collection was representative and varied the cohort comprised people with different health problems, ages and genders [15]. The system was able to handle a variety of health profiles thanks to this varied data collecting which also produced an extensive set of data for in-depth research. Creating a system that could produce holistic health insights independent of the user's history or particular health conditions was the main goal of gathering such a diverse range of data. Following collection, the data was carefully cleaned to remove any values that were missing, outliers and discrepancies that could have affected the analysis. Missing values may occur in real-world datasets such as this one because of tracking gaps, device battery failure or user removal of the wristwatch. Depending on the kind as well as distribution of the values that are not present several imputation approaches were used to handle missing data. When health measures such as blood pressure or heart rate were absent the user's prior measurements or the cohort average for that metric were used to impute the missing values. Using statistical techniques like Z-scores outliers that can skew the study [16] were identified and based on their significance either eliminated or substituted.

Normalizing the dataset [17] was another step in data preparation that made sure all the characteristics were on the same size. It's crucial to normalize health measures such as blood pressure, heart rate and stress levels before using any machine learning approaches because they might vary widely. The Standard Scaler which converts data by removing the mean and normalizing to unit variance was utilized for this purpose. Because machine learning methods like clustering may be sensitive to variations in the magnitude of input variables this phase was essential. In the absence of normalization and the clustering algorithm may overweight characteristics with wider numerical ranges producing skewed results. In order to comprehend the framework and connections within the dataset, data exploration or EDA, was carried out. In order to find trends, anomalies and possible linkages the EDA [18] process started by displaying the geographic distribution of each healthcare measure. To see how heart rate data was distributed among various age groups, genders and activity levels for example histograms, box plots and line charts were used. In a similar vein data on blood pressure was examined for anomalies or patterns such as elevated readings in elderly individuals or at stressful times. It was easier to see typical times of increased stress like throughout the workday or the evening when stress levels were visualized throughout the day. The following grouping procedure and the analysis of the findings were led by the insightful information these visualizations offered.

Correlations between several health measures were investigated during EDA in order to determine the ways in which these factors interacted. For instance, a significant relationship between heart rate and stress levels was found especially during times of intense mental or physical exertion. These relationships were shown using heatmaps which showed that psychological stress tended to rise at the same time as blood pressure and heart rate surges. The association between age and specific health metrics [19] was also investigated using scatter plots which revealed that older people often had higher blood pressure and a lesser amount of physical activity both of which may be significant determinants of their general health. These results laid the groundwork for creating more individualized health advice based on user information.

Covariance is a statistical measure that illustrates the amount that the two factors change at the same time. It assesses the relationship between two variables deviations from their specific means. Positive covariance indicates that if one of the components is above the average of its standard deviation the other is probably above its mean as well and vice opposite for negative variance. However, because covariance lacks a defined measure it is challenging to gauge the strength of the link. It is commonly used in conjunction with the correlation coefficient of analysis to give a deeper comprehension of the linear relationship among variables. This made it possible to spot reoccurring trends such higher stress levels over the week or faster heart rates at particular times of the day. The algorithm was able to offer more timely and precise health suggestions that were customized to the person's daily activity by comprehending these temporal patterns. Following the completion of EDA and a full understanding of the data, the dataset was divided into sets for training and testing. This division was necessary to guarantee that machines learning algorithms could be successfully trained and yet assessed using data that had not yet been viewed. Usually, 80% of the data was set aside to be used for instruction and 20% for testing resulting in an 80/20 split. This split offered a trustworthy way to assess the model performance and helped guarantee that they could generalize effectively to new data. In order to improve the models and avoid overfitting cross-validation procedures were also employed throughout the training phase. This ensured that the program could provide new users with correct insights.

# METHODOLOGY

The methodology of the work is based on an approach to gathering, examining and evaluating medical information from watches used by a variety of individuals. The first step is to collect data from a range of health factors such as blood pressure, heart rate, stress levels and physical activity. Instantaneous information into the user's general health state is provided by this data which is collected over time. Following collection, a thorough data preparation pipeline prepares the raw data for analysis. This pipeline deals with problems that might skew the analysis such irregular data entry, outliers and missing numbers. To ensure that the dataset is accurate and reliable for subsequent stages, it is cleaned using techniques like impute and outlier removal. In particular, Standard Scalers are used to combine all of the health metrics onto a single scale in order to set up the data for effective clustering and pattern identification normalizing approaches. After the data has been processed and uniform it is then clustered using unsupervised methods of machine learning. Dividing the users into distinct groups based on similarities in their medical indicators is the goal. The Elbow Method which calculates the ideal number of subgroups by studying the within-cluster mean of the squares is used to break the data into five distinct categories using the K-Means Clustering method. The clustering technique makes it easier to identify patterns and trends in the data that may not be readily apparent from a first look.

Multiple health variables are used to build the clusters which provide a deeper knowledge of the interactions and variations among users of various health measures. Agglomerative Clustering which offers a hierarchical method to segmentation is also used on the same dataset to validate the categorization findings. A thorough assessment of the grouping process is made possible by this comparison of clustering algorithms which identifies the method that best reflects the subtleties of the data. The system then concentrates on deriving significant insights from the groupings following the clustering analysis. This entails interpreting the clusters and the associated health trends using sophisticated data visualization tools. The visualizations which include scatter plots and heatmaps provide a clear picture of how various individuals are categorized according to their health information. Within each cluster these visual aids assist in identifying important health trends, relationships and outliers. Analytics and statistical techniques are also used to investigate the connections between different health variables both inside and throughout the clusters. The approach seeks to give consumers individualized health insights by examining the traits of each cluster assisting them in comprehending their current state of health and directing them toward improved well-being management. The initiative intends to convert unprocessed health data into useful insights using this thorough technique empowering individuals to make knowledgeable decisions regarding their health.

# A. DATA ANALYTICS

Data analytics [20] is the process of examining raw data to identify trends and insights that might inform choices. Many methods including artificial intelligence, mathematical modeling and data mining are employed to examine and extract useful information from massive databases. Data analytics can use algorithms and statistical models to identify patterns, correlations and anomalies in data. This can offer significant insights on a range of subjects including as market trends, customer behavior and operational effectiveness. Businesses may improve performance, lower risks, boost productivity and encourage innovation with the help of these insights. Every business from banking and healthcare to marketing and manufacturing depends on data analytics. It helps businesses achieve their strategic objectives and offers them a competitive edge. The technology can offer crucial information about the relationships between different physiological traits by using a correlation analysis of medical data collected from wristwatch wearers. For example, association analysis may reveal a positive correlation between cardiac variability and levels of physical exercise suggesting that greater levels of activity are associated with improved cardiovascular health.

Additionally, correlation analysis [21] increases the prediction ability of the wristwatch application by identifying predictive indications for certain health outcomes. For instance, if correlation research reveals a strong link between idleness and an increased risk of obesity, the system can alert users beforehand about the need of maintaining a physically active life to lower the risk of obesity. The development of tailored health treatments that meet the needs of certain customers may also benefit from correlation analysis. A value of -1 indicates a perfect negative linear relationship a value of 0 indicates no linear relationship and a value of 1 indicates a perfect linear positive relationship. Simply stated a correlation coefficient of about 1 indicates that when one variable increases the other usually does too whereas a number close to -1 indicates that the second variable often decreases when the first one rises. This measure is helpful in data analysis for understanding how changes in a single factor may affect another and for identifying correlations between variables. Data analytics is essential for giving people's daily health status insights. These insights can be found in a variety of areas including the distribution of daily smartphone usage time, body temperature, calories burned, blood volume, respiration rate, heart rate, diastolic and systolic blood pressure, stress level, physical activity level, sleep duration and weight. Examining how people use their devices is the first step in determining their daily smartphone usage time (Fig. 1). This can include details about physical activity, screen time patterns and possible effects on overall health. By tracking usage patterns over time, researchers can identify trends, peak usage periods and correlations with other health indicators. This makes it possible to create specific treatments meant to encourage digital health and reduce too much screen usage.

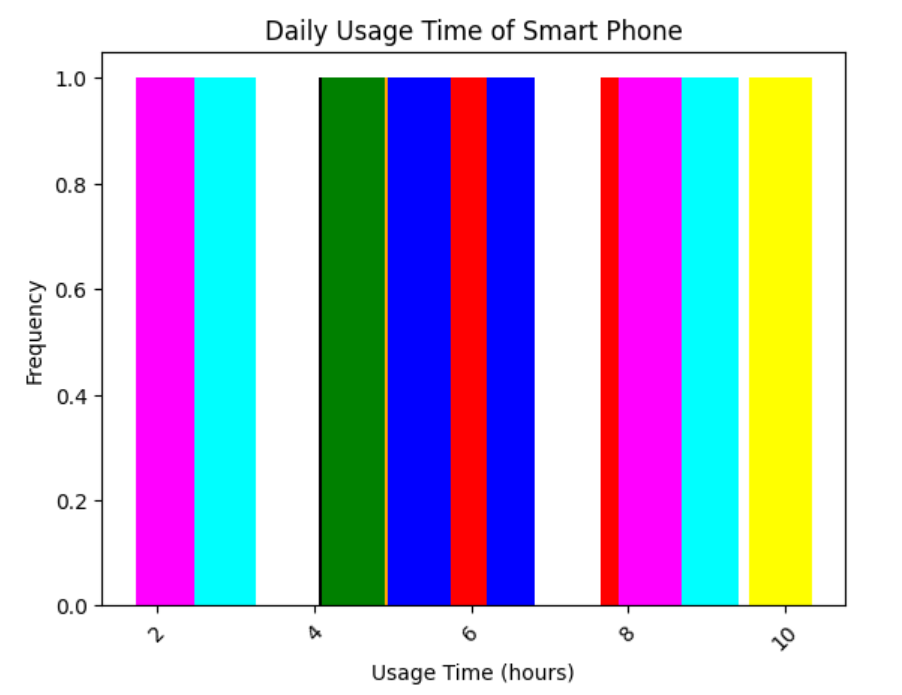


Fig.1 Smart Phone Daily Usage Time

Monitoring the distribution of water intake offers important information about the level of hydration and maintaining the directed intake limits. Researchers can find trends in water intake evaluate how well people follow hydration guidelines and identify dehydration problems by looking at the amount and how often people drink. Individual hydration methods and treatments that increase the intake of water and maintain proper hydration levels can be developed using this data as a guidance. The calorie analysis Burned Distributions could provide details about a person level of physical activity and energy usage. Measuring the inhalation Rate Distribution provides information about a person's physical and respiratory health. Researchers can evaluate respiratory efficiency, spot deviations or problems and spot possible respiratory illnesses or breathing problems by tracking respiration rates all through time (Fig. 2).

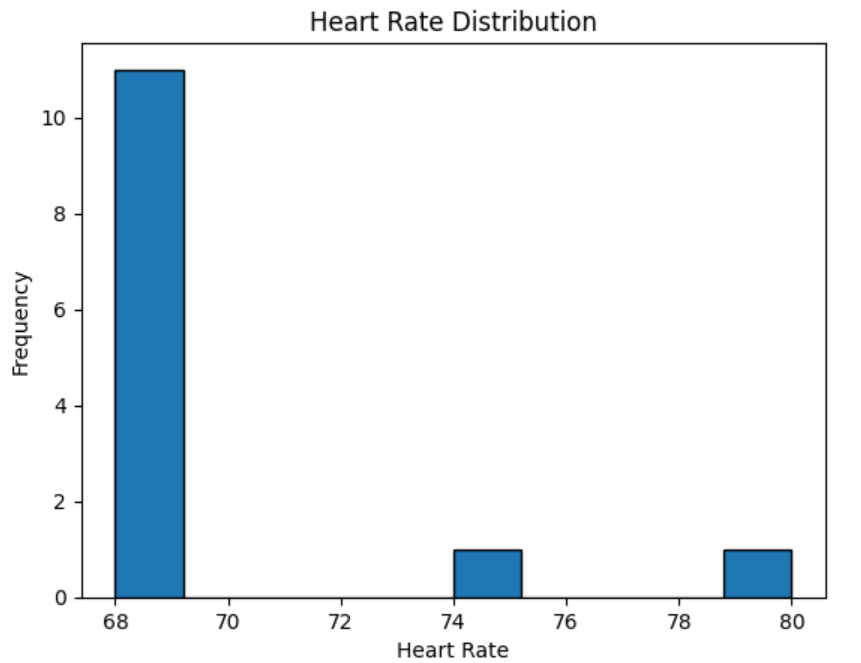


Fig.2 Heart Rate Distribution

Additionally, looking at the Stress Level Range provides insights into people's psychological well-being and stress management. By monitoring variations in levels of stress over time researchers may assess stressors, identify triggers and monitor stress reactions. Programs that promote relaxation techniques, enhance mental health and wellbeing and effectively manage stress might be guided by this. Users sleeping habits and level of peace can be inferred by looking at the Sleep Duration Graph. Researchers can evaluate the quality of sleep, identify sleep disorders or problems and identify factors influencing the state of sleep by monitoring changes in sleep duration over time.

B. CLUSTERING TECHNIQUES

K-Means One of the most well-liked and extensively applied unsupervised machine learning techniques for dividing data into discrete clusters is clustering [22]. It works especially well when the objective is to find naturally occurring groups in data which is important when examining health indicators like blood pressure, heart rate and stress levels. A dataset is divided into 'k' clusters using the K-Means technique and each data point is assigned to the grouping with the closest mean. After initializing 'k' the centroids of cluster at random the method iteratively modifies these the centroids as it allocates data points to the closest cluster according to Euclidean distance. K-Means simplicity and speed are two of its main features which make it appropriate for big datasets.

The Elbow Approach is frequently used to calculate the ideal number of groupings in K-Means. Graphing the within-cluster sum of squares (WCSS) as an estimate of the number of clusters allows the Elbow Method to determine the "elbow" or the point at which the pace of decline in WCSS changes dramatically. In order to balance underfitting and overfitting it recommends the optimal number of clusters. Three clusters (named as Normal, Mild and Severe) were found to be ideal for the health indicators dataset in this work using the Elbow Method (Fig.3). By preventing K-Means from over- or under-segmenting the data this technique produces a more relevant user grouping according to health attributes. K-Means might be used to find unique patterns in users' medical information, such comparable blood pressure readings, heart rate ranges or stress patterns. K-Means has several drawbacks despite its efficacy.

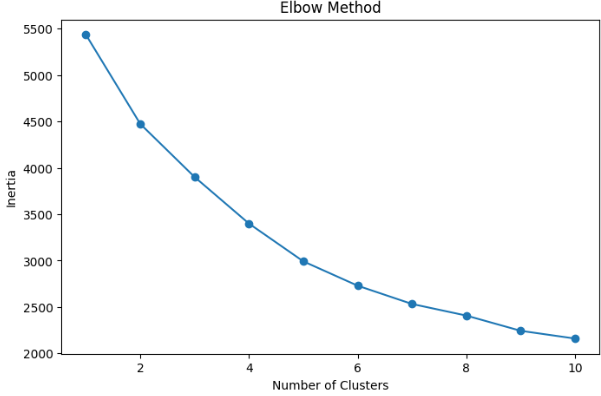


Fig.3 No. of Clusters vs Inertia

Furthermore, the scale of the variables that are input may have an impact on the distance measures used by K-Means such as Euclidean distance. To ensure that every feature participates equally to the cluster process data normalization such as using Standard Scaler is essential when employing K-Means. If the ideal number of groups is uncertain K-Means requirement that the number of clusters be predetermined is another drawback. However, K-Means is a potent technique for revealing hidden patterns in huge health datasets when used appropriately. Agglomerative Clustering on the other hand, provides a hierarchical method of clustering that by combining or dividing clusters at each stage creates a layered hierarchy of clusters. Agglomerative Clustering is more flexible in some situations than K-Means since it does not need the number of groups to be predetermined. Each data point is first treated as a separate cluster by the algorithm which then progressively combines the closest cluster pairings until either all of the data points have been combined into only one cluster or a certain number of groups is achieved. This approach is referred regarded as a "bottom-up" strategy. Agglomerative clustering [23] does however provide a unique set of difficulties. Its computational complexity is one of its primary drawbacks particularly when dealing with big datasets.

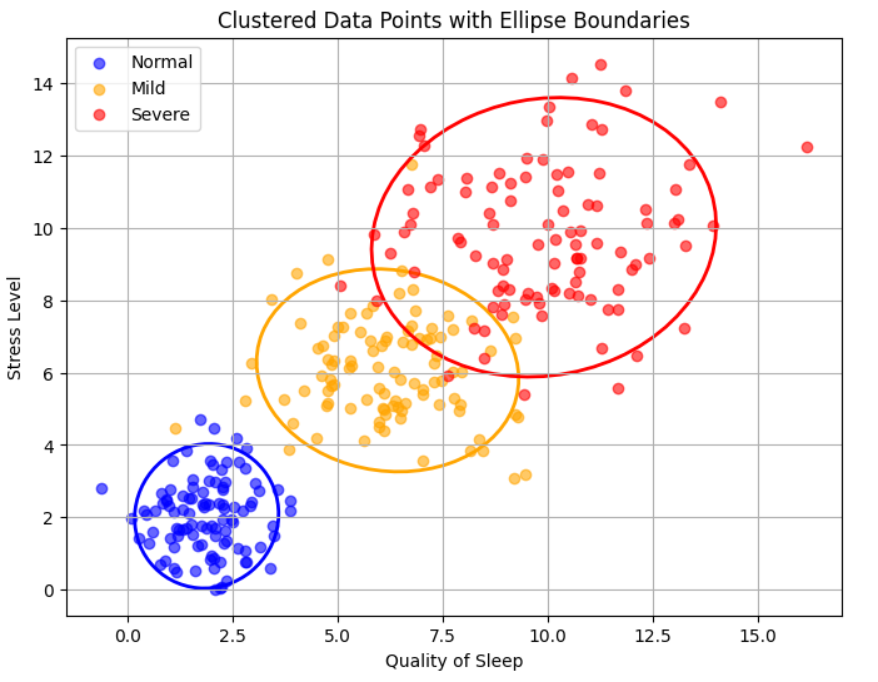


Fig.4 Clustered Data Points of Users

When dealing with hundreds of data points as is frequently the case with medical information gathered from smartwatches the method can be computationally costly and inefficient since it needs to calculate the distance between every conceivable pair of information elements at each step. Because of this it is less appropriate for huge datasets unless the clustering process is approximated using effective approaches. Furthermore, the final clusters might be greatly impacted by the linking criterion selection [24]. Cluster topologies can result from a variety of links and choosing the best one frequently calls for trial and error or domain expertise. In the setting of this work a comparison of the two clustering approaches revealed that K-Means produced more relevant and accurate groups for the specified health parameters. The Elbow Method [25] verified that K-Means worked best with three clusters and the generated clusters were well defined and in line with the analysis's goals. Because of the dataset's complexity and size agglomerative clustering which provides a more thorough and hierarchical picture of the data did not yield as definitive conclusions in this instance. K-Means was more suited for classifying people according to their general health profiles since it could form discrete, flat clusters which enabled the system to provide customized health insights and suggestions for each group.

RESULTS

The outcomes show how well the suggested technique works to clean valuable health information from the data gathered by smartwatches. The information set was divided into clusters using K-means and Agglomerative Clustering following the use of data pretreatment procedures which included managing missing values, scale the data using Standard Scaler and converting it for analysis. Finding unique patterns in customers health data was the main objective in order to provide individualized insights based on variables like stress levels, blood pressure and heart rate. For this dataset the K-Means clustering method optimized using the Elbow Method proved to be the most successful, yielding three ideal clusters silhouette score with 0.85. More individualized health advice was made possible by the distinct health profiles that each cluster represented. Agglomerative clustering produced less distinct groups than K-Means clustering. When the findings were shown the clusters showed clear patterns in important health indicators. One cluster (Fig.4) for example showed persistently high stress levels and heart rates suggesting that the individuals may be at risk for cardiovascular problems. Another cluster indicated a group who would need treatment for hypertension since they had higher blood pressure but reasonable heart rates and stress.

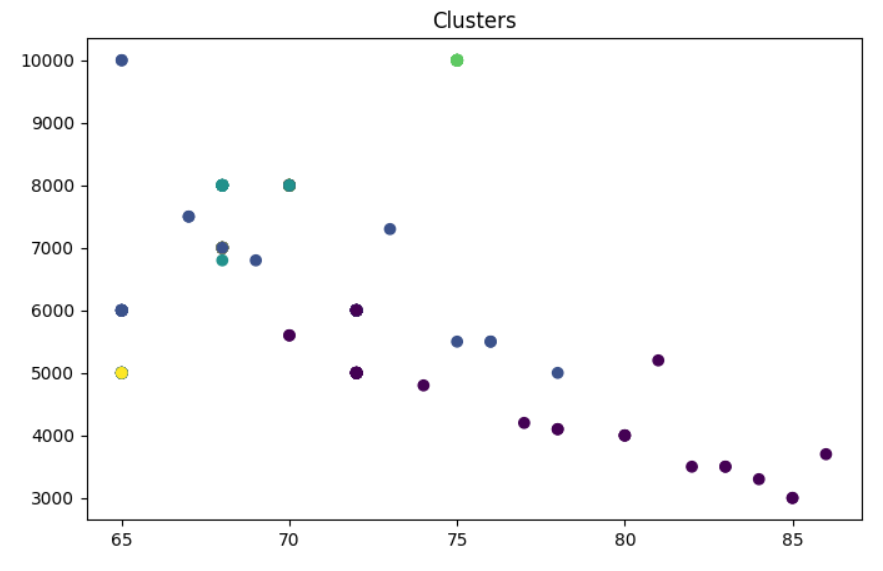


Fig.5 Clusters vs Heart Rate Distribution

Finding these groupings made it possible to derive useful conclusions that opened the door to preventative health measures. Metrics like the Silhouette Score were used to validate the clusters demonstrating that K-Means generated coherent high-quality groups that indicated that the individuals in each cluster had comparable health features. The results showed that K-Means was more successful in separating the data into different health profiles than Agglomerative Clustering. The big complicated dataset of continuous health measures proved more difficult for Agglomerative Clustering to manage despite the fact that it provided a hierarchical representation of the data. Because of the overlapping or ill-defined clusters created by the hierarchical clustering it was difficult to distinguish between various user groups. Furthermore, Agglomerative Clustering was less feasible for this research due to its computing expense which included calculating the distances between every pair of data points. K-Means, on the other hand was the method of choice for assessing the medical information in this study since it offered quicker processing times and better interpretable findings.

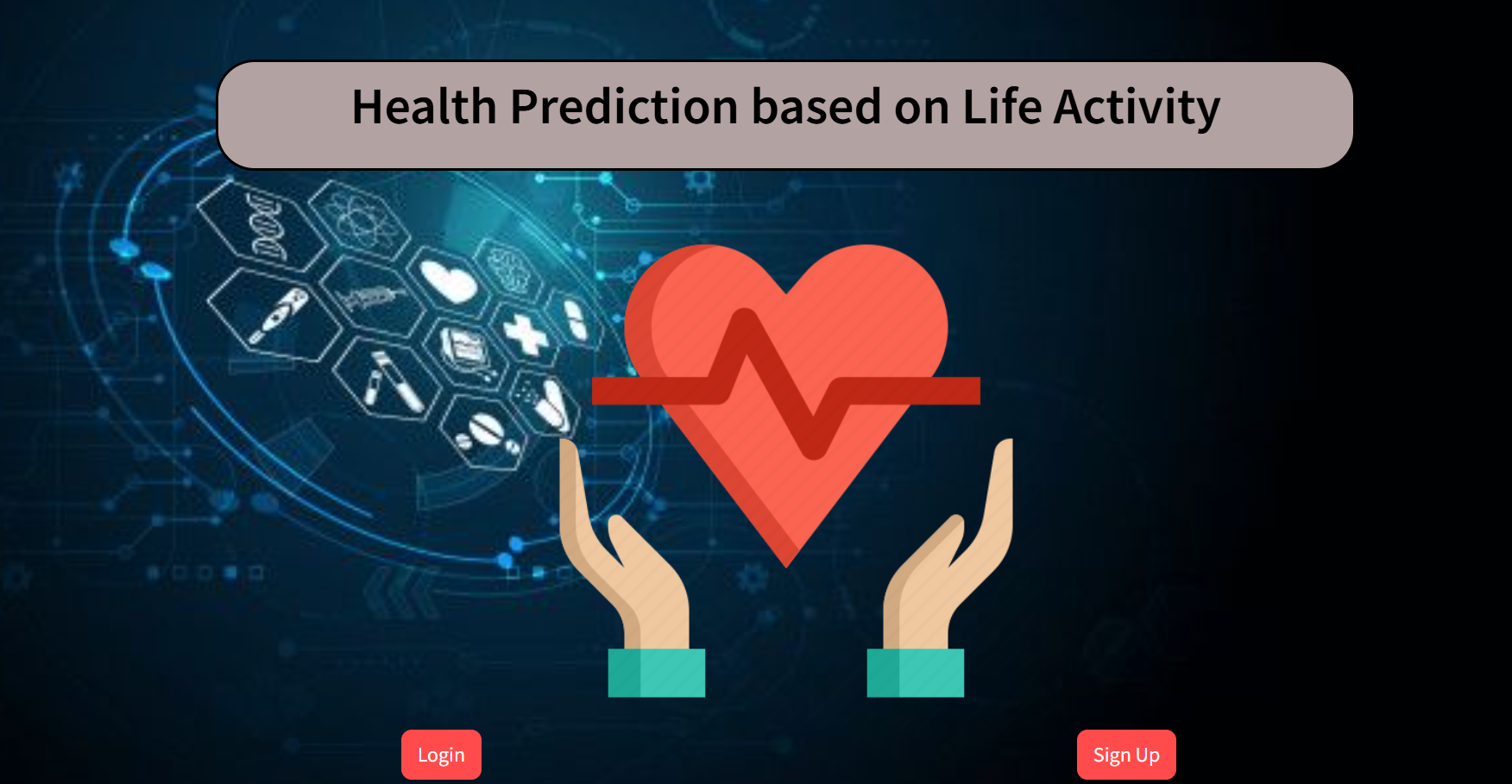


Fig.6 Home Page

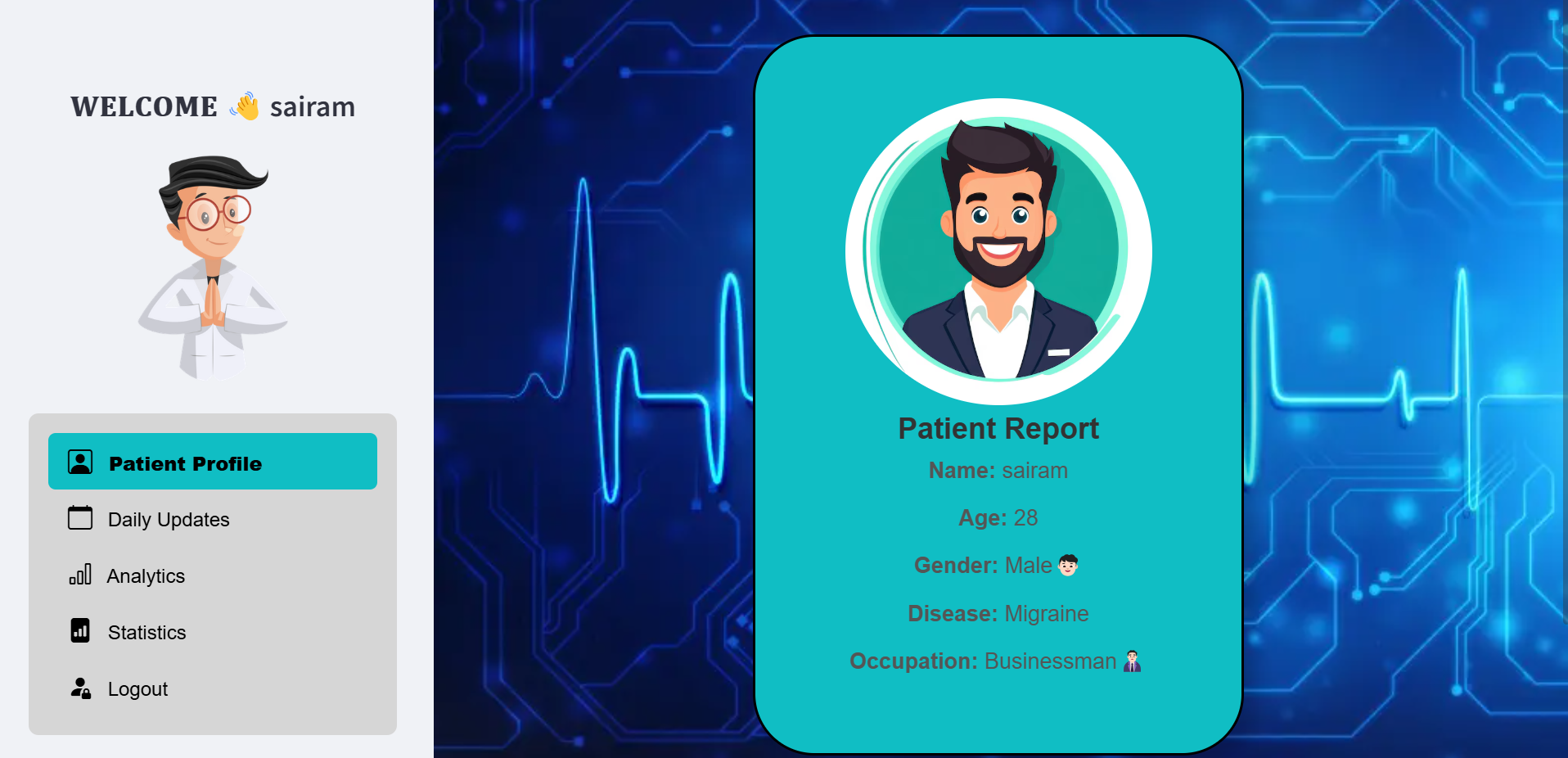


Fig.7 User Profile Page

Subsequent cluster analysis provided valuable information about lifestyle trends and possible health hazards for various user groups. Users in one cluster for instance showed lower levels of stress and consistently had slower heart rates and normal blood pressure indicating that they either lead healthy lifestyles or practice stress management. Conversely those in a different cluster with high blood pressure and higher stress levels had signs of an active lifestyle or elevated levels of tension at work. These findings show how the system may offer users tailored health advice according to their cluster membership promoting healthy lifestyle adjustments. The data visualization tools were essential to understanding the findings. Plotting clusters of data and displaying the relationships between various health parameters allowed the system to clearly and easily convey complicated patterns.

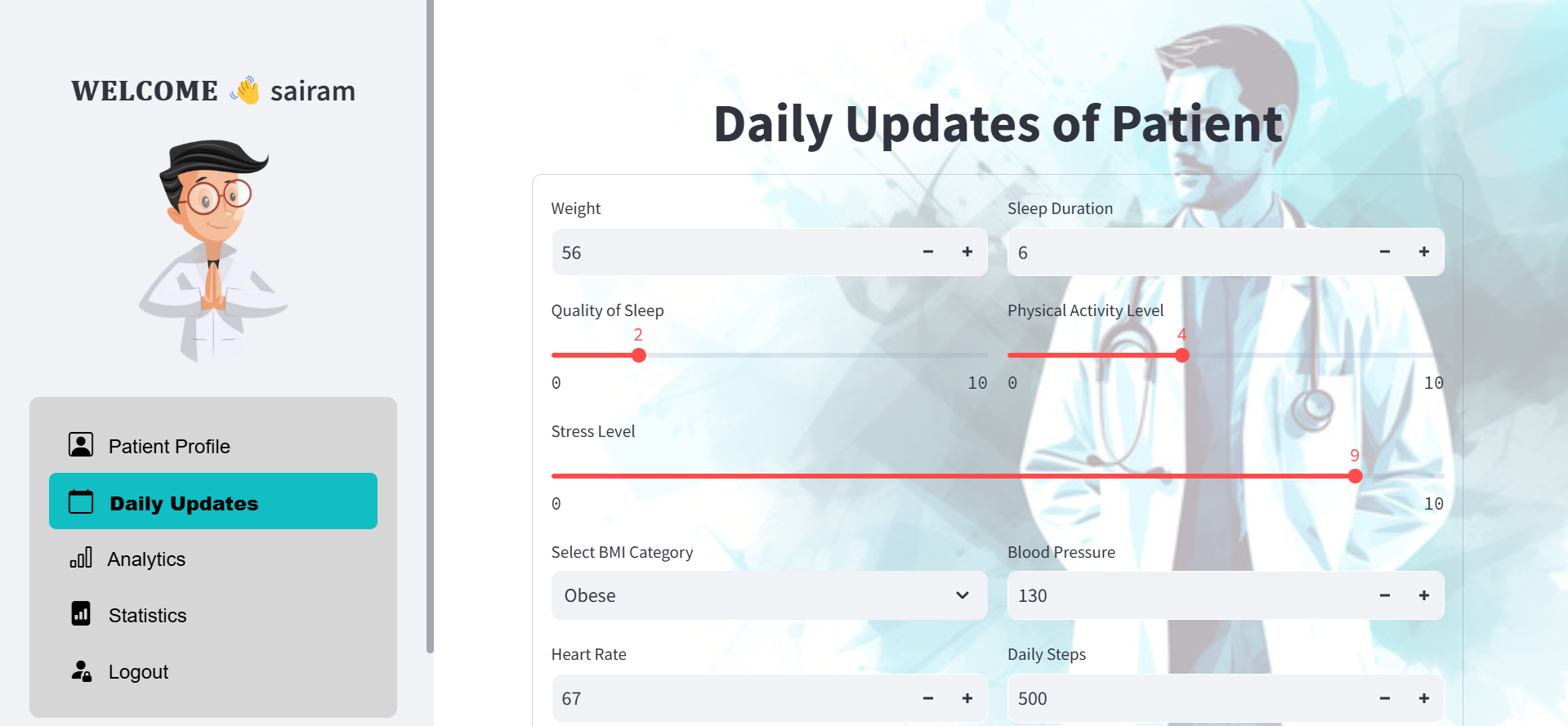


Fig.8 Daily Updates Page



Fig.9 Data Analytics Page

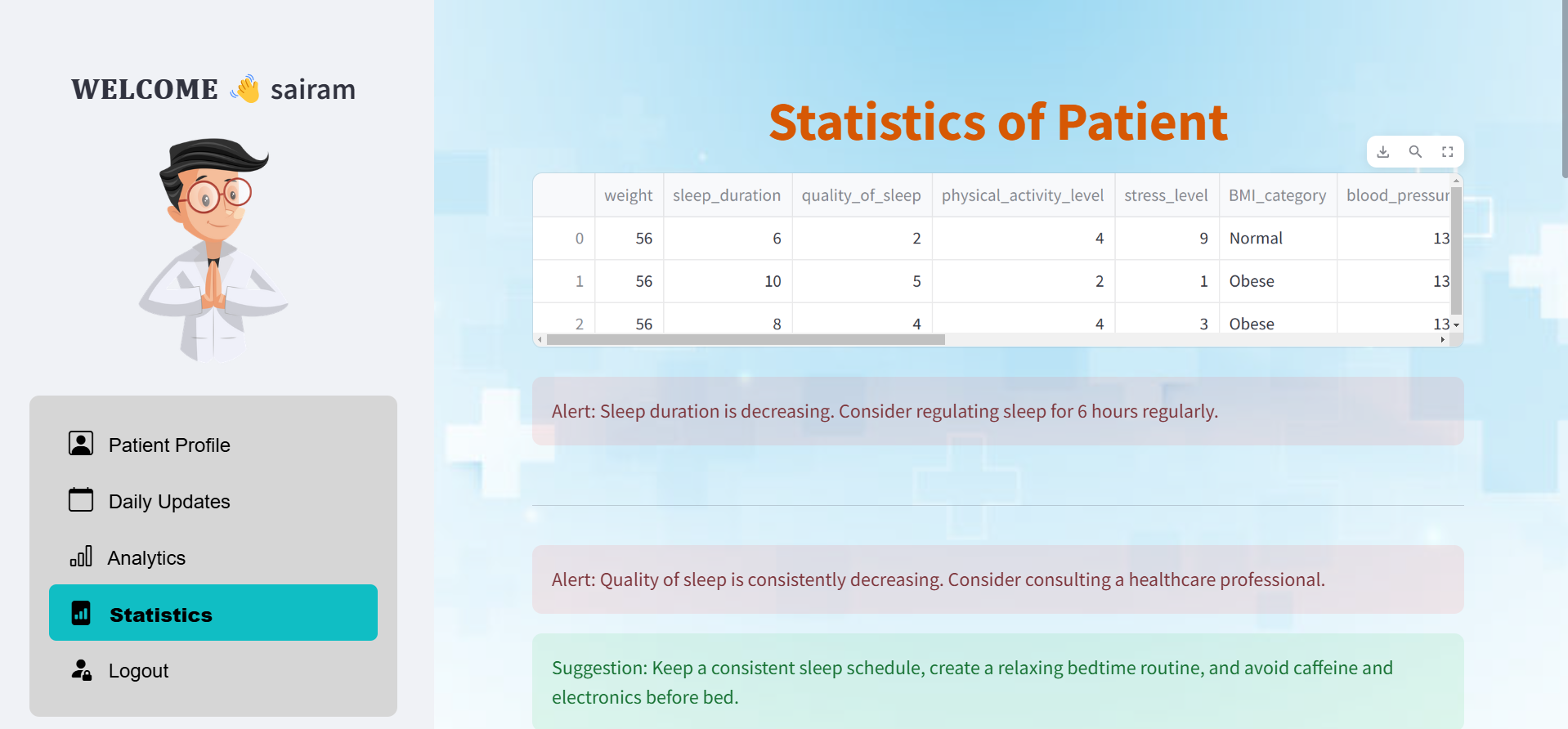


Fig.10 Statistics Page

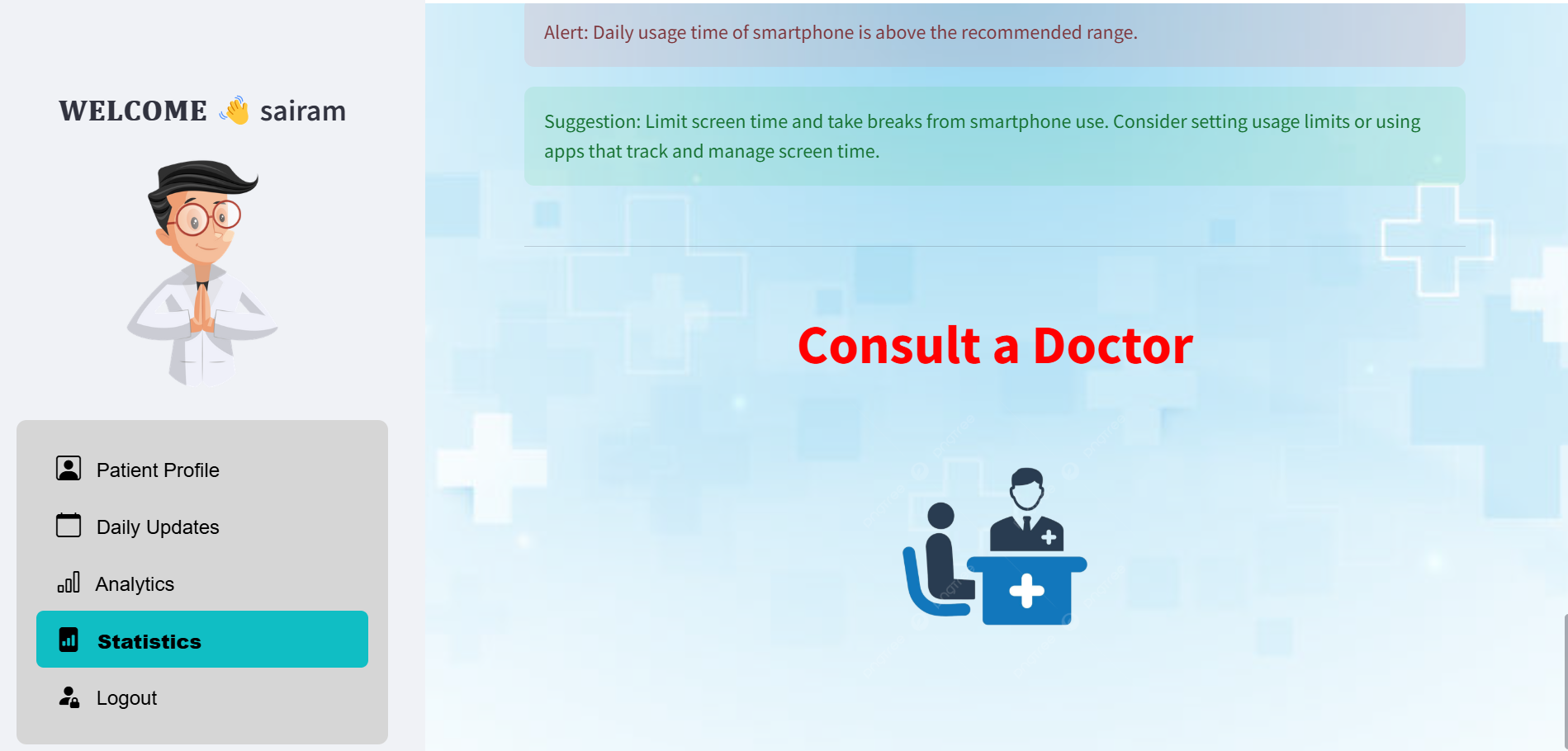


Fig.11 Cluster Grouping of Data

The geographic distribution of health variables within each cluster was shown using heatmaps which highlighted the most prevalent blood pressure, heart rate and stress levels. In addition to making the data simpler to understand these visualizations gave consumers a better understanding of their own health patterns. In general, the results indicate the potential of using algorithms for clustering for wearable health tracking. By efficiently classifying users into applicable clusters based on their health data the system provides specific medical data that can help people take preventative measures to manage their well-being. Both approaches have their benefits but a comparison of K-Means and Agglomerative Clustering shows that K-Means is more appropriate for massive health datasets due to its speed, clarity and capacity to create unique clusters. The knowledge gathered from this study has the potential to transform health monitoring by assisting individuals in better understanding their health and making informed decisions regarding preventative care.

##### CONCLUSION

By indicating the importance of using clustering algorithms for evaluating health metrics obtained from smartwatch data this study provides an integrated approach to managing and tracking health. Using K-Means clustering and advanced data preparation techniques the study identified various user groups with unique health profiles based on important parameters including blood pressure, heart rate and stress levels. By dividing the population into various groups it is possible to provide tailored health insights and focused suggestions encouraging a proactive strategy to healthcare that enables people to make knowledgeable lifestyle choices. The results highlight how wearable technology may be used to better understand and manage human health in addition to tracking health parameters. Additionally, the comparison between K-Means and Agglomerative Segmentation showed that K-Means was the better approach for this dataset providing findings that are easy to understand and crucial for practical insights. This study highlights the value of data-driven strategies in advancing preventative healthcare in addition to adding to the expanding area of digital health monitoring.

##### References

1. Powers, Rob, et al. "Smartwatch inertial sensors continuously monitor real-world motor fluctuations in Parkinson’s disease." Science translational medicine 13.579 (2021): eabd7865.
2. Dubey, Vinita, et al. "Improving preventive service delivery at adult complete health check-ups: the Preventive health Evidence-based Recommendation Form (PERFORM) cluster randomized controlled trial." BMC family practice 7 (2006): 1-12.
3. Phillips, Shane A., et al. "Advances in health technology use and implementation in the era of healthy living: implications for precision medicine." Progress in cardiovascular diseases 62.1 (2019): 44-49.
4. Danon-Hersch, Nadia, et al. "Prevalence, awareness, treatment and control of high blood pressure in a Swiss city general population: the CoLaus study." European
5. Journal of Preventive Cardiology 16.1 (2009): 66-72.
6. Dias, Duarte, and João Paulo Silva Cunha. "Wearable health devices—vital sign monitoring, systems and technologies." Sensors 18.8 (2018): 2414.
7. Prabhakaran, D., et al. "Differences in the prevalence of metabolic syndrome in urban and rural India: a problem of urbanization." Chronic Illness 3.1 (2007): 8-19.
8. Prabhu, Tanishkka Senthil, Aradhana Balodi Bhardwaj, and Vinod Kumar Shukla. "Influence of Smart Fitness Devices and Mobile Applications on Perceived Stress Levels of Adults: A United Arab Emirates-Based Study." Industry 4.0 and People Analytics. Apple Academic Press 205-235.
9. Kansal, Tushar, et al. "Customer segmentation using K-means clustering." 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS). IEEE, 2018.
10. Soto-Vergel, Angelo J., et al. "Transforming ground disaster response: Recent technological advances, challenges, and future trends for rapid and accurate real-world applications of survivor detection." International Journal of Disaster Risk Reduction 98 (2023): 104094.
11. Mohammad, Gouse Baig, et al. "Prediction of health disease based on day-to-day life activity using Machine Learning Approach." (2021).
12. Thakur, Saurabh Singh, Pradeep Poddar, and Ram Babu Roy. "Real-time prediction of smoking activity using machine learning based multi-class classification model." Multimedia tools and applications 81.10 (2022): 14529-14551.
13. Guzik, Przemyslaw, and Marek Malik. "ECG by mobile technologies." Journal of electrocardiology 49.6 (2016): 894-901.
14. Meng, Yiwen, et al. "A machine learning approach to classifying self-reported health status in a cohort of patients with heart disease using activity tracker data." IEEE journal of biomedical and health informatics 24.3 (2019): 878-884.
15. Thayer, Julian F., et al. "A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health." Neuroscience & Biobehavioral Reviews 36.2 (2012): 747-756.
16. Szklo, Moyses. "Population-based cohort studies." Epidemiologic reviews 20.1 (1998): 81-90.
17. Seo, Songwon. A review and comparison of methods for detecting outliers in univariate data sets. Diss. University of Pittsburgh, 2006.
18. Singh, Dalwinder, and Birmohan Singh. "Investigating the impact of data normalization on classification performance." Applied Soft Computing 97 (2020): 105524.
19. Wang, Shuai. "A semi-arid climate's use of exploratory data analysis (EDA) as a reliable non-parametric method for geochemical mapping." Environmental Research 261 (2024): 119654.
20. Boddapati, Mohan Sai Dinesh, et al. "Creating a Protected Virtual Learning Space: A Comprehensive Strategy for Security and User Experience in Online Education." International Conference on Cognitive Computing and Cyber Physical Systems. Cham: Springer Nature Switzerland, 2023.
21. Tukey, John W. "The future of data analysis." Breakthroughs in Statistics: Methodology and Distribution. New York, NY: Springer New York, 1962. 408-452.
22. Gogtay, Nithya J., and Urmila M. Thatte. "Principles of correlation analysis." Journal of the Association of Physicians of India 65.3 (2017): 78-81.
23. Qtaish, Amjad, et al. "Optimization of K-means clustering method using hybrid capuchin search algorithm." The Journal of Supercomputing 80.2 (2024): 1728-1787.
24. Ackermann, Marcel R., et al. "Analysis of agglomerative clustering." Algorithmica 69 (2014): 184-215.
25. Bholowalia, Purnima, and Arvind Kumar. "EBK-means: A clustering technique based on elbow method and k-means in WSN." International Journal of Computer Applications 105.9 (2014).