28. Clustering Patient Profiles Using K-Means: Analyzing Blood Pressure and Cholesterol Levels for Enhanced Healthcare Decision-Making

**Abstract**

This study presents the application of the K-means clustering algorithm to categorize patient data based on two critical health indicators: blood pressure (mmHg) and cholesterol levels (mg/dL). The dataset is divided into five clusters, each representing distinct patient groups with similar health characteristics. The visualization reveals meaningful patterns in the clustering results, which may be leveraged to identify patient risk profiles and improve personalized healthcare strategies. The findings demonstrate the utility of unsupervised machine learning in analyzing complex medical datasets and highlight opportunities for enhancing clinical decision-making.

**Introduction**

Clustering is a fundamental unsupervised machine learning technique that groups data points based on their similarities. In the medical field, clustering methods like K-means are widely used to identify patterns in patient data that may be indicative of health conditions or risk factors. This paper utilizes the K-means clustering algorithm to analyze a dataset consisting of blood pressure and cholesterol measurements from a large cohort of patients. Our goal is to segment the data into clusters representing different health profiles, which can assist in risk assessment, treatment planning, and patient management.

**Methods**

**Dataset Overview**

The dataset contains two primary variables:

1. **Blood Pressure (mmHg):** A continuous variable representing the systolic blood pressure of patients.
2. **Cholesterol Levels (mg/dL):** A continuous variable indicating the total cholesterol levels in the blood.

**Clustering Approach**

1. **K-means Clustering Algorithm:**
   * The K-means algorithm partitions the data into five clusters, each represented by a centroid, which is the mean of all data points within that cluster. The number of clusters (k = 5) was chosen to capture diverse patient profiles while maintaining interpretability.
2. **Visualization:**
   * A scatter plot illustrates the distribution of patient data points across the two variables, with different colors indicating the five clusters. The black crosses represent the centroids of each cluster, serving as prototypes for their respective classes.
3. **Cluster Validation:**
   * The silhouette method and within-cluster sum of squares (WCSS) were used to evaluate the quality of clustering, ensuring that the selected number of clusters adequately represents the data's inherent structure.

**Results**

**Clustering Visualization**

The scatter plot shows a clear segmentation of the data into five clusters:

1. **Cluster 1 (Red):** Represents patients with lower blood pressure (100–120 mmHg) and moderate cholesterol levels (150–200 mg/dL).
2. **Cluster 2 (Green):** Comprises patients with slightly higher blood pressure (120–140 mmHg) and a wide range of cholesterol levels (200–250 mg/dL).
3. **Cluster 3 (Cyan):** Includes patients with relatively high blood pressure (140–160 mmHg) and high cholesterol levels (250–300 mg/dL).
4. **Cluster 4 (Blue):** Consists of patients with the highest blood pressure readings (160–180 mmHg) and moderate to high cholesterol levels (250–300 mg/dL).
5. **Cluster 5 (Pink):** Contains patients with moderate blood pressure (100–160 mmHg) and lower cholesterol levels (100–150 mg/dL).

**Cluster Centroids and Distribution**

* The centroids, marked by black crosses, represent the central tendency of each cluster, providing a prototype for the group characteristics.
* The even distribution of clusters across the plot suggests that the dataset includes a diverse range of patient profiles. The distinct separation between clusters indicates that the variables used (blood pressure and cholesterol levels) are effective in differentiating patient groups.

**Discussion**

The K-means clustering results reveal distinct patient groups based on blood pressure and cholesterol levels, offering insights into potential health risks and treatment strategies:

1. **Cluster Implications:**
   * **Cluster 1 and 5 (Lower Risk):** Patients in these clusters exhibit lower to moderate blood pressure and cholesterol levels, suggesting a lower cardiovascular risk profile. These groups may require standard monitoring and lifestyle interventions.
   * **Clusters 2, 3, and 4 (Higher Risk):** These clusters represent patients with higher blood pressure and cholesterol levels, which are known risk factors for cardiovascular diseases. Tailored interventions, including medication and regular monitoring, may be necessary for these groups.
2. **Clinical Utility:**
   * The clustering model can aid healthcare professionals in identifying patient subgroups that may benefit from targeted interventions, thereby enhancing personalized medicine. The identified clusters provide a foundation for developing risk prediction models and creating decision-support tools for clinical practice.
3. **Limitations and Future Work:**
   * The analysis is based solely on two variables, and additional factors such as age, gender, smoking status, and genetic predisposition could refine the clustering results. Future studies could incorporate these variables to create more comprehensive patient profiles.

**Conclusion**

This study demonstrates the effectiveness of K-means clustering in categorizing patients based on blood pressure and cholesterol levels. The resulting clusters provide valuable insights into different health profiles, facilitating targeted interventions and improving patient outcomes. The findings underscore the potential of unsupervised learning techniques in healthcare data analysis and encourage further exploration into more complex models that consider additional patient characteristics.

**References**

* Include relevant literature on K-means clustering, patient data analysis, and personalized medicine to support the findings and discussions in the paper.

This research provides a foundation for future studies leveraging clustering techniques to enhance patient care and healthcare decision-making.