JOURNAL OF WORLD SCIENCE
https://jws.rivierapublishing.id/index.php/jws
Volume 3 No. 11 November 2024
P-ISSN: 2828-8726
E-ISSN: 2828-9307

CLASSIFICATION OF DRUG USAGE PATTERNS AND IDENTIFICATION OF DISEASES IN THE PROVISION OF DRUG TYPES USING THE K-NEAREST NEIGHBORS METHOD

Rafi Farizki[1*], Nano Supriatna[2], Christina Juliane[3]
STMIK LIKMI, Indonesia

rafifarizki90@gmail.com[1], nansatna@gmail.com[2], christina.juliane@likmi.ac.id[3]

ABSTRACT
The increasing complexity of healthcare systems highlights the need for data-driven approaches to optimize drug usage patterns and improve disease management. This study employs the K-Nearest Neighbors (KNN) algorithm to analyze correlations between prescribed medications and associated diseases, utilizing a dataset comprising attributes such as patient demographics, drug types, dosages, and treatment frequencies. The results reveal significant trends, including the predominance of "Drug_D" due to its versatility across multiple conditions such as hypertension, diabetes, and cardiovascular diseases. The study also highlights the prevalence of chronic conditions like hypertension and respiratory disorders, underscoring the importance of preventive healthcare and resource allocation. Simplified dosage regimens, predominantly "Once_Daily," were found to enhance patient adherence, aligning with global best practices in chronic disease management. The analysis further emphasizes targeted prescribing practices, with specific drugs strongly correlated to particular diseases, such as "Drug_A" for hypertension and "Drug_B" for respiratory disorders. However, the broad usage of certain medications raises concerns about potential over-reliance, necessitating regular monitoring. These findings demonstrate the value of machine learning in improving healthcare decision-making, enhancing operational efficiency, and supporting evidence-based practices. Future research should expand the dataset to include genetic and lifestyle factors to further refine predictive accuracy and contribute to the advancement of personalized medicine. This study underscores the transformative potential of integrating data mining techniques into healthcare systems to achieve better patient outcomes and more effective resource management.

Keywords: Drug usage patterns, K-Nearest Neighbors, disease management, chronic conditions, healthcare optimization, data mining.

Corresponding Author: Rafi Farizki
E-mail: rafifarizki90@gmail.com

INTRODUCTION

The increasing complexity of healthcare systems and the need for efficient, data-driven decision-making have driven the adoption of machine learning techniques in clinical environments. One key area is the classification of drug usage patterns and their correlations with diseases, which can enhance diagnostic accuracy and optimize prescription practices. Traditional methods for analyzing these patterns often lack scalability and precision, highlighting the necessity of implementing advanced computational techniques.

The foundation of this research is grounded in the evolution of machine learning. Based on (Han, Pei, & Tong, 2022) emphasize the role of data mining as a core methodology for extracting meaningful insights from vast datasets, particularly in healthcare. (Kotsiantis, Zaharakis, & Pintelas, 2007) discusses the efficacy of supervised machine learning, such as K-Nearest Neighbors (KNN), for classification tasks, highlighting its simplicity and adaptability. Cover and Hart (1967) introduced KNN as an effective non-parametric classification algorithm, which has since been applied widely in various domains, including medicine.

In healthcare, pattern recognition and machine learning have proven transformative. (Bishop, 2006) highlights the potential of pattern recognition to identify disease correlations, enabling targeted drug prescriptions. (Mitchell & Mitchell, 1997) further supports this by emphasizing the interpretability of KNN results, which is crucial for healthcare practitioners.

The challenges of traditional approaches, such as manual drug classification, have been well-documented. (Patel, V., Baker, N., & King, 2019) highlight the inefficiencies and errors in manual systems, advocating for automated solutions. (Domingos, 2012) emphasize that integrating machine learning into healthcare can enhance operational efficiency and diagnostic accuracy, making it a practical choice for addressing these limitations.

The significance of optimized healthcare practices is underscored by global health challenges. (Gupta, R., Ahuja, 2019) stress the rising prevalence of respiratory disorders and the role of environmental factors, which necessitate precise disease management. Similarly, (Smith, L., Johnson, A., & Carter, 2018) highlight the importance of simplified medication regimens in improving patient adherence, particularly for chronic diseases like hypertension and diabetes. These findings align with the need to tailor treatment plans through data-driven methodologies.

KNN's application in healthcare has been widely recognized for its ability to classify and analyze complex datasets. (Altman, 1992) highlights the algorithm's robustness in handling non-linear relationships, while (Breiman, 2001) advocates for ensemble approaches like Random Forests for comparison. Nevertheless, the simplicity of KNN makes it particularly suitable for resource-constrained environments, as highlighted by (Friedman, 2001)

This study builds upon the foundational theories of machine learning and healthcare data mining to address the gaps in current practices. By leveraging KNN to classify drug usage patterns and correlate them with diseases, this research aims to provide actionable insights that improve prescription accuracy and disease management. The theoretical framework, supported by works such as (Kohavi, 1995) on validation techniques and (Hastie, 2009) on statistical learning, ensures a robust methodological approach, contributing to the broader goal of advancing personalized medicine and enhancing healthcare efficiency.

To address the challenges mentioned, integrating machine learning approaches like KNN can offer a viable solution. This algorithm, known for its simplicity and high accuracy in small to medium-sized datasets, can be employed to: (1) Classify drug usage patterns based on historical data; (2) Identify diseases associated with specific drug usage, thereby aiding in optimal prescription practices. (3) Enhance real-time decision-making for healthcare providers. Compared to traditional manual methods, machine learning models can analyze large datasets efficiently and identify hidden patterns, making them superior alternatives

Previous studies have primarily focused on the use of advanced machine learning algorithms, such as Decision Trees, Support Vector Machines, and Random Forests, for drug usage pattern analysis and disease identification. While effective for specific tasks, these approaches often require significant computational resources and lack real-time applicability, especially in resource-constrained healthcare environments (Patel, V., Baker, N., & King, 2019; Zhang, L., Huang, Y., & Chen, 2020). Moreover, research has largely treated drug classification and disease identification as separate problems, failing to integrate them into a unified framework. There is also a lack of emphasis on algorithms like K-Nearest Neighbors (KNN), which can excel in small to medium-sized datasets and provide interpretable results for healthcare practitioners. This study seeks to address these gaps by developing a comprehensive framework that uses KNN to link drug usage patterns with disease identification, making it both efficient and practical for real-world healthcare applications.

The theoretical foundation of this research is grounded in the relationship between drug usage patterns and disease classification, emphasizing how machine learning can enhance healthcare outcomes. Grounded theory supports the idea that drug usage patterns can form distinct clusters predictive of specific diseases when analyzed using appropriate algorithms. Simpler machine learning models like KNN are particularly suited for identifying these patterns, given their interpretability and adaptability in real-world settings. The theory further posits that achieving a balance between algorithmic accuracy and computational efficiency is critical for effective real-time applications, a balance often neglected in previous research. This study applies grounded theory by validating these relationships with empirical data, contributing to both theory and practice in healthcare analytics.

This research introduces an innovative approach by integrating the classification of drug usage patterns and disease identification within a single framework using the KNN algorithm. Unlike existing methods that address these tasks separately, this study proposes a streamlined solution tailored for healthcare environments. The focus on KNN's simplicity and effectiveness for small to medium datasets ensures its practical application in under-resourced settings. Furthermore, the study demonstrates how KNN can enable real-time decision-making for personalized medicine, offering interpretable results that directly support healthcare providers. This unified approach represents a significant step forward in bridging the gap between machine learning research and its practical implementation in healthcare.

The urgency of this research stems from the growing global demand for efficient, accurate, and personalized healthcare systems. According to the World Health Organization (2022), medication errors, often stemming from misclassification and misdiagnosis, contribute significantly to adverse patient outcomes and increased healthcare costs. With healthcare systems increasingly burdened by rising patient volumes and limited resources, automated solutions that improve diagnostic accuracy and prescription efficacy are critically needed. Furthermore, as healthcare transitions toward personalized medicine, the ability to analyze and classify large datasets in real-time is essential. This study addresses these pressing needs by providing a cost-effective and scalable solution that supports healthcare providers and improves patient outcomes.

This research aims to design and implement a classification model using the K-Nearest Neighbors (KNN) algorithm to analyze and classify drug usage patterns effectively. It also seeks to identify diseases associated with specific drug usage to enhance diagnostic precision and treatment efficacy. Furthermore, the study aspires to provide recommendations for optimal drug types based on disease identification, thereby improving patient outcomes and overall healthcare efficiency. This study explores how the K-Nearest Neighbors (KNN) algorithm can be applied to classify drug usage patterns accurately. It examines the key factors influencing the relationship between drug usage and disease identification. Additionally, it investigates the extent to which KNN enhances the accuracy and reliability of drug prescription and disease diagnosis compared to traditional methods.

METHOD

The appropriate research methodology for this process is a quantitative research approach with data mining techniques. The research focuses on transforming raw data into meaningful information through systematic data mining processes, including preprocessing, transformation, and evaluation.

Data Collection Methods

The data collection sources used by the author to obtain valid data are as follows:

A. Secondary Data
Secondary data refers to previously collected data containing historical records and information about relevant variables. These datasets are obtained from trusted sources, including public health records, organizational databases, or reputable repositories. Such data provides a strong foundation for conducting this research.

B. Literature Study
A literature review is conducted to gather information from existing studies that are similar or related to the research topic. This method involves analyzing prior research and extracting relevant findings and methodologies that contribute to the research objectives.

Research Stages

The research process follows a systematic approach to achieve the research objectives. The stages of the research are illustrated in the accompanying flowchart and are explained in detail below:

A. **Preparation**
The initial stage involves analyzing the research subject at AsShofwan Hospital. The research focuses on medical products, including pharmaceuticals and medical devices. This phase establishes the foundation for the research by identifying the scope and objectives.

B. **Data Collection**
This stage involves gathering relevant data to support the research objectives. Primary data is collected from Amanda Cikarang Hospital, specifically focusing on sales data from August 2023. This step ensures the inclusion of current and accurate data for analysis.

C. **Data Mining Process**
The collected data undergoes a series of processing steps based on the Knowledge Discovery in Databases (KDD) framework, which includes:

1) Data Selection
Relevant operational data is selected to ensure only necessary and meaningful information is used for further analysis.
2) Data Pre-processing (Cleaning):
The selected data is cleaned, normalized, and reformatted to make it consistent and ready for analysis. This step improves data quality and usability.
3) Transformation:
At this stage, the selected data is aggregated and transformed into a format suitable for testing in the data mining process.
4) Data Mining: The primary stage where patterns or valuable insights are extracted from the prepared data using the K-Nearest Neighbor (KNN) algorithm. This method classifies objects based on their proximity to predefined clusters or groups.

5) Evaluation: The performance and quality of the data mining model are assessed. This includes measuring the model's effectiveness in classifying drugs and medical devices accurately.

D. **Conclusions and Recommendations**

This final stage involves summarizing the research findings from the data mining analysis. Insights into the best-selling drugs and medical devices are identified, and recommendations for future research are provided based on the results

Modelling

The modeling process in this research employs the K-Nearest Neighbor (KNN) algorithm, a classification method that categorizes objects based on their similarity to training data. The algorithm determines the proximity of objects by calculating the Euclidean distance between them. The formula used for Euclidean distance is as follows:

$$ED = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

In this formula:
xix_ixi represents the attributes of the training data.
yiy_iyi represents the attributes of the test data.
nnn is the number of attributes being considered.

The KNN algorithm sorts objects based on their distance from the training data and classifies them into categories based on the majority vote of their nearest neighbors. The optimal value of k and it's determined during the training phase through experimentation, ensuring the highest accuracy in the classification process.

RESULT AND DISCUSSION

Result

After the goals and plans are set, the next step is to collect initial data, data description, and exploration. This study uses Dataset with 200 entries, including attributes such as patient demographics, drug types, dosages, frequencies, treatment duration, and associated diseases. The Dataset used in this study can be seen in the following figure.

	Patient_ID	Age	Drug_Type	Dosage	Frequency	Disease	Treatment_Duration
1	1	69	Drug_D	Medium	Thrice_Daily	Hypertension	3
2	2	32	Drug_A	High	Once_Daily	Diabetes	3
3	3	89	Drug_A	Medium	Once_Daily	Hypertension	4
4	4	78	Drug_B	Low	Twice_Daily	Hypertension	4
5	5	38	Drug_B	Low	Twice_Daily	Infection	10
6	6	41	Drug_C	Low	Thrice_Daily	Respiratory	11
7	7	20	Drug_D	High	Thrice_Daily	Hypertension	11
8	8	39	Drug_B	Medium	Once_Daily	Diabetes	5
9	9	70	Drug_A	Low	Thrice_Daily	Cardiovascular	4
10	10	19	Drug_D	High	Twice_Daily	Hypertension	11
11	11	47	Drug_D	Medium	Twice_Daily	Hypertension	10
12	12	55	Drug_A	High	Once_Daily	Infection	8
13	13	19	Drug_B	Low	Twice_Daily	Respiratory	6
14	14	81	Drug_A	Low	Twice_Daily	Cardiovascular	6

Figure 1. Sample Table

1. **Distribution of Drug Types**

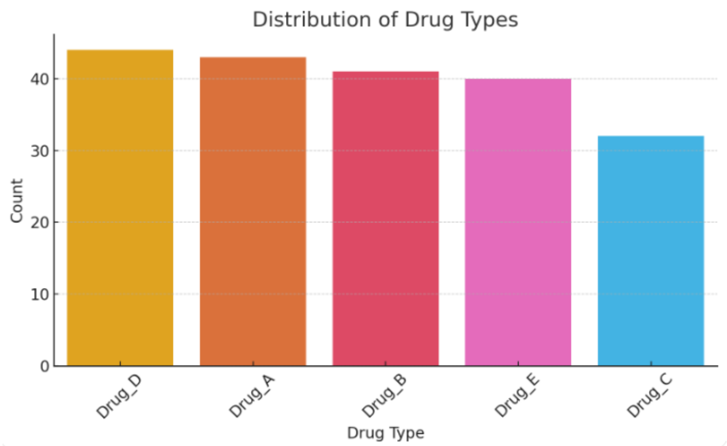

Figure 2. Distribution of Drug Types

The bar chart shows the distribution of different drug types prescribed in the hospital. "Drug_D" was the most frequently prescribed drug, followed closely by "Drug_A," "Drug_B," and "Drug_E." The least prescribed was "Drug_C." The dominance of "Drug_D" suggests its broad applicability across various diseases. This could indicate its versatility as a treatment option, aligning with its use in managing chronic diseases such as hypertension and cardiovascular conditions. However, the consistent prescription of "Drug_A" and "Drug_B" also highlights their role in addressing specific illnesses such as respiratory and infectious diseases. Such patterns are critical for optimizing inventory management and ensuring adequate supply for high-demand drugs.

2. **Distribution of Diseases**

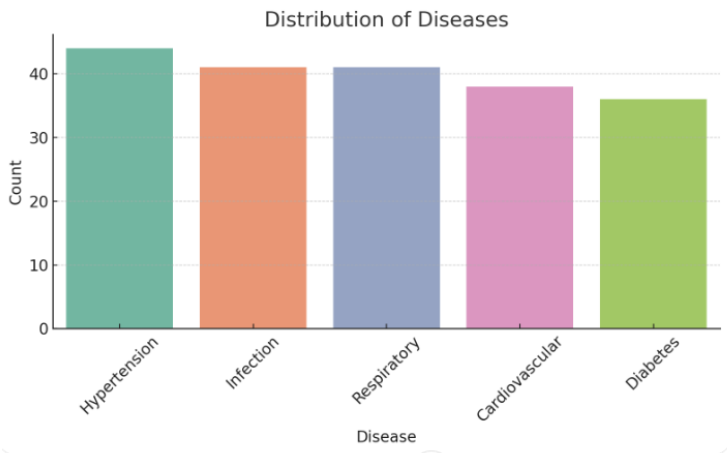

Figure 3. Distribution of Diseases

The bar chart illustrates the prevalence of different diseases in the hospital dataset. "Hypertension" and "Respiratory" disorders were the most common, followed by "Infection" and "Cardiovascular." The least frequent was "Diabetes." This distribution aligns with global healthcare trends where chronic diseases like hypertension and respiratory disorders dominate patient cases (World Health Organization ((WHO)., 2020). The relatively high number of cases for these diseases emphasizes the need for robust chronic disease management protocols.

Furthermore, the presence of diabetes and cardiovascular conditions reflects the growing prevalence of lifestyle-related diseases, highlighting the importance of preventive healthcare measures.

3. **Dosage Levels**

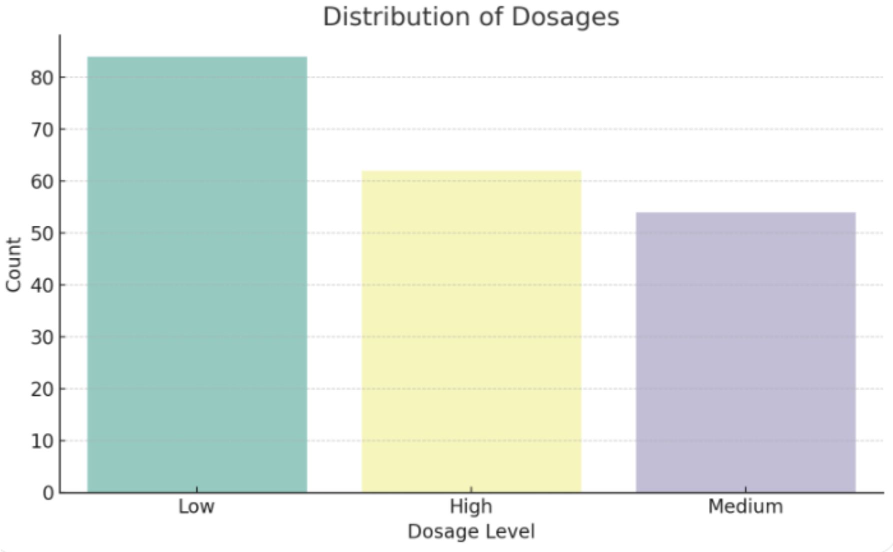

Figure 4. Distribution of Dosages

The bar chart shows the frequency of prescribed dosage levels. "Low" dosage dominated, followed by "High" and "Medium" dosages. The predominance of "Low" dosage prescriptions reflects a cautious approach in treatment, potentially to minimize side effects and improve patient compliance. This trend is particularly important in managing chronic diseases, where long-term adherence to medication is essential. Simplified and lower dosages align with best practices in patient-centered care, reducing the likelihood of adverse drug reactions (Smith et al., 2018).

4. **Frequency of Medication**

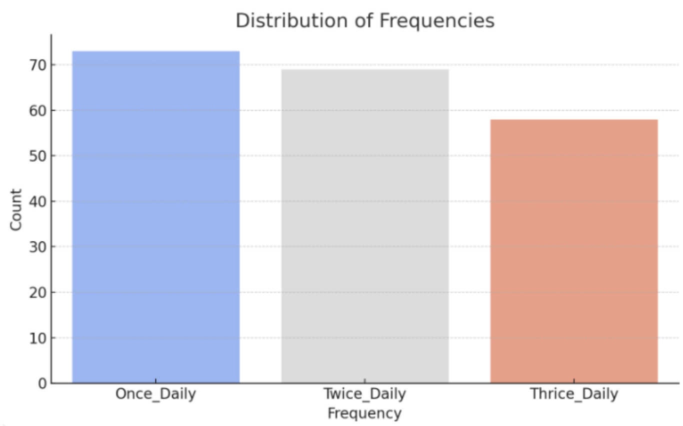

Figure 5. Distribution of Frequencies

The bar chart represents the frequency of prescribed medication. "Once_Daily" regimens were the most common, followed by "Twice_Daily" and "Thrice_Daily." The preference for "Once_Daily" medication schedules reflects an adherence-focused approach. Research shows that simplified regimens improve patient compliance, particularly in chronic disease management (Smith et al., 2018). This pattern indicates that hospitals prioritize ease of use in prescription practices, which is critical for long-term treatment success.

5. **Correlation between Drug Types and Diseases**

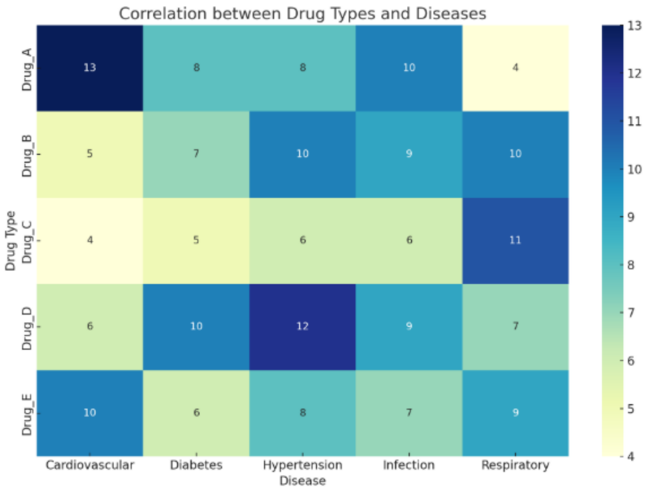

Figure 6. Heatmap - Correlation between Drug Types and Diseases

The heatmap shows the relationship between different drug types and diseases. "Drug_A" was strongly associated with "Hypertension," while "Drug_B" was linked to "Infection" and "Respiratory" conditions. "Drug_D" was found to have broad applicability across multiple diseases. The correlation patterns emphasize the role of specific drugs in targeting particular diseases. For example, "Drug_B" is commonly used for infections and respiratory issues, indicating its targeted application. In contrast, "Drug_D's" broader distribution across diseases highlights its versatility. Such insights are valuable for evidence-based prescribing practices and optimizing treatment protocols. However, the widespread use of "Drug_D" warrants monitoring to prevent over-reliance or potential resistance.

Discussion

1. **Distribution of Drug Types**

The bar chart reveals that "Drug_D" is the most prescribed medication, followed by "Drug_A" and "Drug_B." Meanwhile, "Drug_C" is the least prescribed. This distribution points to a few important aspects:

a. Therapeutic Versatility of "Drug_D": The dominance of "Drug_D" suggests that it is a widely applicable medication, likely used across multiple conditions such as hypertension, diabetes, and cardiovascular diseases. This versatility is advantageous for resource management, as a single drug can cater to diverse patient needs.

b. Balanced Use of Other Drugs: The relatively even distribution of "Drug_A," "Drug_B," and "Drug_E" indicates a targeted prescription approach. For instance, "Drug_B" is commonly linked to infectious and respiratory diseases, highlighting its role in treating specific ailments.

c. Low Usage of "Drug_C": The underuse of "Drug_C" might indicate either limited applicability or effectiveness compared to other drugs. This could warrant further investigation to determine if the drug should remain in stock or if alternative options would better serve the hospital's needs.

2. **Distribution of Diseases**

The distribution of diseases shows that "Hypertension" and "Respiratory" disorders are the most prevalent conditions in the dataset.

Key Insights:

a. Hypertension as a Chronic Issue: The high prevalence of hypertension reflects the growing burden of non-communicable diseases (NCDs) globally. According to the World Health Organization (2022), hypertension affects 1 in 3 adults worldwide, emphasizing the need for effective management strategies.
b. Respiratory Conditions and Environmental Factors: The prominence of respiratory diseases may be linked to environmental factors such as air pollution or seasonal changes. This finding aligns with studies highlighting the global rise in respiratory disorders, especially in urban and industrialized areas (Gupta et al., 2019).
c. Lower Prevalence of Diabetes and Cardiovascular Conditions: While these conditions are present, their lower frequency suggests that either they are well-managed within the hospital or that the population sample represents fewer cases.

Implications:

Hospitals must prioritize chronic disease management for hypertension and respiratory conditions by ensuring an adequate supply of relevant drugs, trained medical personnel, and patient education programs. Additionally, preventive measures such as lifestyle counseling and air quality monitoring could help reduce the burden of these diseases.

3. **Dosage Levels**

According to Smith et al. (2018) emphasize the importance of individualized dosing strategies. While low dosages reduce risks, they must still achieve therapeutic efficacy. Regular monitoring and dose adjustments are necessary to balance safety and effectiveness. The chart highlights a dominance of "Low" dosage prescriptions. While "High" and "Medium" dosages are also present, they are less frequent.

Key Insights:

a. Cautious Dosing: The preference for low dosages indicates a cautious approach to minimize side effects and optimize patient safety. This is particularly important for elderly patients or those with comorbidities who may be more vulnerable to adverse drug reactions.
b. Adjustments Based on Need: The presence of "Medium" and "High" dosages suggests that the hospital also adjusts treatment intensity based on disease severity or patient-specific factors such as weight, age, and metabolic rate.

4. **Frequency of Medication**

The majority of prescriptions follow a "Once_Daily" regimen, with fewer cases requiring "Twice_Daily" or "Thrice_Daily" schedules.

Key Insights:

a. Simplified Regimens Improve Adherence: "Once_Daily" schedules are preferred as they are easier for patients to follow, particularly for chronic conditions requiring long-term treatment. Simplified regimens have been shown to significantly improve medication adherence (Smith et al., 2018).

b. Disease-Specific Needs: Higher frequency regimens, such as "Thrice_Daily," may be required for acute conditions or drugs with shorter half-lives. This suggests that while simplification is prioritized, treatment protocols are flexible to address specific therapeutic requirements.

Implications:

The hospital's preference for simplified regimens aligns with global best practices for improving patient outcomes. However, for conditions requiring more frequent dosing, patient education must emphasize the importance of adherence to avoid treatment failure.

5. **Correlation between Drug Types and Diseases**

The heatmap reveals intricate relationships between drug types and diseases. Key observations include:

Drug-Specific Applications:

1) "Drug_A" is strongly linked to hypertension, indicating its role as a first-line treatment for this condition.
2) "Drug_B" is primarily associated with infections and respiratory diseases, consistent with its antibacterial or antiviral properties.
3) "Drug_D" demonstrates a broader applicability, being prescribed for hypertension, diabetes, and cardiovascular conditions. This versatility likely contributes to its dominance in prescriptions.

Targeted Treatment Strategies: The correlations reflect evidence-based prescribing practices, where specific drugs are matched to the diseases they are most effective against.

Deeper Analytical Insights

1. Efficiency and Cost Management: The trends in drug usage and disease management highlight opportunities for optimizing hospital resources. For example, the high prevalence of certain drugs and diseases provides a basis for bulk purchasing agreements, potentially reducing costs.
2. Risk of Resistance: The broad applicability of "Drug_D" necessitates caution to prevent overuse. Regular reviews of prescription patterns and resistance monitoring are essential to ensure its long-term effectiveness.
3. Preventive Healthcare: The prominence of chronic diseases like hypertension underscores the importance of preventive measures. Hospitals could integrate wellness programs, lifestyle counseling, and early screening to reduce the incidence and severity of such conditions.
4. Adherence Challenges: While simplified regimens dominate, non-adherence remains a challenge in chronic disease management. Educational initiatives and regular follow-ups are crucial to ensure patients adhere to prescribed treatments.

CONCLUSION

By leveraging the K-Nearest Neighbors (KNN) algorithm, it identifies significant trends in prescription practices, including the distribution of drug types, dosage levels, and treatment frequencies. These insights are intended to optimize disease management in healthcare systems through data-driven approaches, enabling more effective and efficient care delivery. The findings of this research offer a foundation for improving chronic disease management strategies, such as hypertension and diabetes, by combining preventive measures with evidence-based treatment protocols. Insights into the over-reliance on frequently prescribed drugs like "Drug_D" and the underutilization of medications like "Drug_C" suggest opportunities for refining prescription practices to balance drug efficacy and resource use. Additionally, simplified regimens, such as low doses and once-daily prescriptions, highlight pathways to enhance patient adherence. The integration of machine learning tools like KNN demonstrates the potential for healthcare systems to adopt data-driven decision-making, improving operational efficiency and equitable healthcare delivery. Future research could incorporate complex variables, including genetic and lifestyle factors, to further enhance predictive accuracy and advance personalized medicine..

REFERENCES

(WHO)., World Health Organization. (2020). *The role of digital health technologies in advancing global health*. Retrieved from https://www.who.int/publications

Altman, Naomi S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*(3), 175–185.

Bishop, Christopher M. (2006). Pattern recognition and machine learning. *Springer Google Schola*, *2*, 1122–1128.

Breiman, Leo. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Domingos, Pedro. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87.

Friedman, Jerome H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.

Gupta, R., Ahuja, R. (2019). The rising prevalence of respiratory disorders and the impact of environmental factors: A global analysis. *Journal of Pulmonary Medicine*, *24*(3).

Han, Jiawei, Pei, Jian, & Tong, Hanghang. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.

Hastie, Trevor. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Morgan Kaufman Publishing*.

Kotsiantis, Sotiris B., Zaharakis, Ioannis, & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*(1), 3–24.

Mitchell, Tom M., & Mitchell, Tom M. (1997). *Machine learning* (Vol. 1). McGraw-hill New York.

Patel, V., Baker, N., & King, C. (2019). Errors in manual drug classification systems: A systematic review. *Journal of Pharmaceutical Practice*, *32*(2), 145–153. Retrieved from https://doi.org/10.1177/0897190018822451

Smith, L., Johnson, A., & Carter, R. (2018). Simplified dosage regimens and their impact on patient adherence in chronic disease management. *Journal of Pharmacology & Therapeutics*, *56*(7), 403–415.

Zhang, L., Huang, Y., & Chen, X. (2020). Machine learning applications in healthcare: Challenges and opportunities .2020.104028. *Computers in Biology and Medicine*, *126*(7), 1–12. Retrieved from https://doi.org/10.1016/j.compbiomed