Clustering Analysis of Physicians' Performance Evaluation: A Comparison of Feature Selection Strategies to Support Medical Decision-Making

Amani Mustafa Ghazzawi[1*], Alaa Omran Almagrabi[2], Hanaa Mohammed Namankani[3]
Department of Information Systems-Faculty of Computing and Information Technology,
King Abdulaziz University, Jeddah., Saudi Arabia[1]
Department of Management Information Systems-Faculty of Business Administration, Taif University, Taif, Saudi Arabia[1]
Department of Information Systems-Faculty of Computing and Information Technology,
King Abdulaziz University, Jeddah, Saudi Arabia[2]
Department of Information Systems-Faculty of Computing and Information Technology,
King Abdulaziz University, Jeddah, Saudi Arabia[3]

***Abstract*—Evaluating physicians' performance is one of the fundamental pillars of improving the quality of healthcare in medical institutions, as it contributes to measuring their ability to provide appropriate treatment, interact effectively with patients, and work within healthcare teams. This study aims to explore the impact of attribute selection on the accuracy of physician clustering using the K-Means algorithm, to improve physician performance assessment. Three datasets containing professional, medical, and administrative attributes were analyzed, such as age, nationality, job title, years of experience, number of operations, and evaluations from various entities. The optimal number of clusters was determined using the Elbow and Silhouette Score methods. The results showed that the original feature set and Lasso features performed best at k = 3, with a clear distinction between clusters. The "three-star" cluster performed well at k = 2 but lost some fine details. It was also shown that attribute selection directly affects the number and accuracy of clusters resulting from clustering, allowing for a clearer classification of physician categories. The study recommends using either original features or Lasso features to achieve more effective clustering, which supports improved recruitment, training, and management decision-making processes in healthcare organizations.**

Keywords—Physicians; performance; evaluation; clustering; k-means; features; decision making

I. Introduction

Physician performance evaluation is essential for improving the quality of healthcare and ensuring the provision of effective and safe medical services. With the advancement of data analysis techniques, it has become possible to use modern methods, such as clustering, to group physicians based on objective criteria based on professional performance, multiple evaluations, and practical experience. The effectiveness of these methods depends largely on the selection of appropriate attributes that help identify differences between different categories of physicians. Several studies have been conducted on physician performance evaluation. Brennan et al. found that most studies in the healthcare sector relied on manual assessment of physician performance through direct management, encompassing both professional and personal aspects. In their traditional approach, he and Baker used attributes such as physician personal information (age, speciality, gender), medical knowledge, communication skills, peer evaluation, patient satisfaction, and practical experience [1]. Kuemerle demonstrated that these methods, despite their high cost, are comprehensive and effective [2]. Zhang's study relied on several tools, including regression analysis, integrating Norman's theory of action and Reson's theory of human error, as well as developing a medical practice framework, a MOC program, systematic searches of electronic databases, a discretionary survey system, Pearson's correlation coefficient, and linear mixed models [3]. In contrast, another study used artificial intelligence to evaluate physician performance. Shi et al. relied on online text consultations between physicians and patients, using Python programming and a simple partitioning ordinal mapping (SVMOP). Model features included the number of medical terms used by the physician, the number of patient questions, as well as predictive features such as tact and emotional words [4].

With the advancement of data analysis technology, it has become possible to use advanced methods such as clustering to identify hidden patterns within complex data sets. Clustering, according to Xu & Wunsch, is a data analysis technique that relies on dividing data into homogeneous groups, each with similar characteristics [5]. This method can be applied to physician data to extract the most influential attributes in evaluating their performance, thereby improving evaluation quality. The importance of this study lies in exploring the extent to which clustering contributes to improving the accuracy of predictive models for evaluating physician performance. In a study conducted by Ghazzawi et al., different datasets collected from an Egyptian hospital were compared. They sought to identify the attributes that best represent physician performance evaluation criteria using regression analysis. The dataset included various attributes, such as nationality, job title, years of experience, and number of surgeries, as well as multiple ratings from different stakeholders, such as patients, nurses, and supervisors. The results showed that the set of attributes was divided into three groups: the original feature set, the 3-star feature set, and the best Lasso features [6]. The current study focuses on analyzing the effect of feature selection on clustering accuracy. Several methods for feature selection are compared to determine the most appropriate methods for grouping physicians according to different performance criteria. Regression methods were applied to these attributes in Ghazzawi et al.'s study to identify the most influential attributes, which this study will use in clustering to test the accuracy of the resulting classifications. This research aims to improve administrative decision-making within healthcare organizations by providing more accurate physician ratings. By analyzing this data, a deeper understanding of the differences between physicians and the factors influencing their performance can be achieved, helping to improve decision-making, evaluation strategies, and professional development in healthcare organizations. This study may also contribute to improving predictive models used to evaluate physicians, enhancing the ability to allocate resources more efficiently, guiding training programs, and designing physician evaluation policies based on accurate and reliable data.

A. Significance of the Study

This study is of great importance in the context of improving physician performance evaluation in healthcare institutions using the clustering method. With the rapid development of technology and big data, it has become necessary to apply advanced data analysis methods such as clustering to better understand the patterns and factors influencing physician performance. This study contributes to:

1) Improving assessment accuracy: By analyzing diverse data and using the clustering method, the most influential attributes in physician performance evaluation are identified, enhancing the accuracy of evaluations and helping improve decision-making.

2) Supporting advanced data analytics in healthcare: The study promotes the use of big data-based data analytics methods, such as clustering, to improve evidence-based healthcare decisions.

3) Discovering hidden patterns: The study helps uncover unseen patterns that may contribute to improving the effectiveness of predictive models, contributing to improving the quality of healthcare.

4) Achieving strategic improvements: The study provides strategic insights for improving training programs, allocating resources, and developing effective policies for evaluating physicians based on the most influential attributes.

B. Objectives

1) Identifying the most important attributes for physician performance evaluation: The study aims to identify the key attributes that should be emphasized when applying clustering to improve performance evaluation.

2) Comparing the impact of clustering on the accuracy of predictive models: The study aims to compare the impact of clustering on improving the accuracy of predictive models for evaluating physician performance and provide recommendations for its use.

3) Analyzing the relationship between various traits and physician performance: The study aims to analyze the relationship between traits extracted from the data (such as years of experience, job title, number of operations) and physician performance in various assessments.

4) Achieving strategic improvements in healthcare institutions: The study aims to provide strategic insights for improving resource allocation, training programs, and policy development that impact physician performance evaluation based on clustering results.

5) Uncovering patterns and factors influencing performance evaluation: The study aims to uncover hidden patterns that may contribute to improved decision-making related to physician performance in healthcare.

C. Research Problem

The study's problem is to identify the most appropriate attributes to focus on when evaluating physician performance using the clustering method. There is an urgent need to understand the relationship between various attributes (such as nationality, job title, years of experience, number of operations, and patient evaluation) and physician performance. Furthermore, the study raises questions about the extent to which the clustering method can improve the accuracy of predictive models and discover effective patterns that contribute to making accurate data-driven healthcare decisions. In this context, a set of questions guides this study:

- What are the most important attributes in evaluating physician performance when applying the clustering method?

This question aims to identify the attributes that primarily contribute to evaluating physician performance after analyzing different datasets. This contributes to improving evaluation accuracy and ensuring that the models used reflect the most influential attributes.

- What are the differences between different datasets (original attributes, 3-star attributes, and the Lasso model) in terms of their impact on physician performance evaluation?

This question aims to compare the impact of the original attributes, the attributes extracted through 3-star regression, and the Lasso predictive model on performance evaluation, and to determine which dataset provides more accurate and reliable results.

- Can using the clustering method reveal new patterns in physician performance evaluations that were not apparent using traditional methods?

This question focuses on exploring the ability of the clustering method to discover new patterns in data that may contribute to improving decisions related to physician performance evaluation, which may not be apparent using traditional methods.

- How can clustering results be used to improve physician training programs and resource allocation within hospitals?

This question aims to examine how clustering results can be applied to improve training programs and resource allocation within hospitals, leading to improved physician performance and the delivery of high-quality healthcare.

- What is the relationship between the attributes extracted from the datasets and physician performance in various assessments (such as the evaluations of patients, nurses, and supervisors)?

This question helps examine the relationship between attributes such as years of experience, number of operations, and physician performance as evaluated by different stakeholders, such as patients, nurses, and supervisors.

Through these questions, the study aims to provide a comprehensive vision on how to improve physician performance assessment using the clustering approach, which enhances the ability to make accurate and reliable decisions based on pivotal data, thus improving overall performance in the health system.

D. Research Contributions

1) Comparison of different datasets: This study contributes by providing a detailed comparison between three datasets collected from a hospital in Egypt, including the original attributes, attributes extracted using 3-star regression, and the Lasso model. This helps identify the most effective attributes for evaluating physician performance.

2) Expanding understanding of clustering in healthcare: The study provides scientific insights by applying clustering to physician performance evaluation, demonstrating how this approach can improve results by identifying the most influential patterns and attributes.

3) Analyzing the relationship between different attributes and performance: By examining the relationship between attributes such as nationality, job title, years of experience, and other attributes, and the varied performance of physicians in different evaluations, this study contributes to providing new insights to support decision-making in healthcare institutions.

4) Enhancing predictive capability in healthcare models: By using advanced methods such as clustering, the study contributes to improving predictive models that can support strategic decisions for physicians and hospital management.

E. Paper Layout

The paper's reminder is organized as follows: in Section II, a Related works is presented; in Section III, the Methodology that includes Datasets Used, Methodology for Determining the Optimal Number of Clusters, Finding the Optimal Number of Clusters, Analysis of clustering results and appropriate decisions, in Section IV, The research work is concluded by expressing direction, in Section V, Study Limitations, in Section VI, Future work, which will open new avenues of exploration and discovery, for upcoming research work.

II. Literature Review

Many previous studies have focused on evaluating the professional performance of physicians using various data analysis techniques, with an emphasis on selecting the most influential attributes in the evaluation process.

Campbell et al, aimed to evaluate the performance of physicians in the United Kingdom through a cross-sectional survey involving patients and colleagues, and the data were analyzed using principal component analysis and regression. The results confirmed that communication skills, clinical competence, and professionalism, along with age, gender, and specialty, play a fundamental role in assessing medical performance [7]. Mirfat et al. confirmed that individual, psychological, and organizational factors play a crucial role in understanding physician performance. Psychological factors were found to have the strongest direct influence, while organizational factors showed a positive but statistically insignificant effect [8]. On the other hand, Cassel et al. demonstrated that intrinsic motivation, such as achievement and patient appreciation, along with extrinsic incentives such as financial rewards and recognition, significantly influence physician motivation levels [9]. In another study, Cola et al. demonstrated that the success of physician-scientists depends on a range of factors, including role balance, autonomy, organizational support, teamwork, mentorship, and the ability to build relationships. These multidimensional factors are essential for understanding and improving physician performance in academic medical settings [10]. Jin et al. also noted that factors influencing healthcare worker performance, such as burnout and anxiety, were analyzed before and after the COVID-19 pandemic, providing deeper insights into improving physician performance [11]. In addition, William et al. identified 22 key variables that influence medical lecturers' performance, most notably leadership, commitment, and credit scores, reflecting the complexity of the interrelationships that govern the performance evaluation process [12]. Overeem et al, used multi-source feedback (MSF) tools to evaluate physicians based on patient ratings, colleague assessments, and the physicians' self-evaluations. The results showed significant differences between the physicians' self-evaluations and the ratings provided by others, indicating the importance of using multi-source data to analyze performance [13]. Bindels et al, a professional performance evaluation system for physicians was implemented in a Dutch medical center through peer conversations based on the principles of appreciative inquiry and continuous feedback. The study emphasized the importance of continuous professional development and periodic feedback in improving physicians' performance [14]. Study by Ho and Baker: The General Medical Council (GMC) has developed a framework for good medical practice, which includes assessing physicians' performance every five years by measuring knowledge, communication skills, decision-making, and patient-centered medical practice [15]. Dias et al, a systematic review of 69 studies addressing the use of machine learning in evaluating physician competence was conducted, analyzing the impact of various features on professional performance using decision trees, support vector machines, and random forests. The study found that the specialties of surgery and radiology were the most affected by these technologies and emphasized that feature selection significantly impacts the accuracy of the models [16]. As Ghazzawi et al. focused on analyzing data from an Egyptian hospital using regression techniques to examine the attributes that most influence physician performance. The datasets included attributes related to physicians, such as age, nationality, job title, years of experience, number of publications, and surgeries, as well as various ratings from patients, nurses, and human resources. The first dataset consists of the original dataset, which contains a wide range of attributes potentially relevant to performance evaluation. The second dataset represents a selection of attributes extracted using regression and receiving a 3-star rating. The third dataset includes the most effective attributes in predictive models, such as the Lasso model. The study focuses on this comparison between the different datasets, aiming to highlight the attributes that most significantly influence the clustering results and physician performance evaluation. The results demonstrate that regression analysis can be an effective tool for healthcare administrators to help reduce medical error rates, providing a framework for data-driven decision-making. Table I summarises the results of all the regression models in the Ghazzawi et al. study, which we will rely on in our current study [6].

TABLE I. IMPORTANCE OF FEATURES IN PREDICTING THE THREE MODELS, ADAPTED FROM STUDY [6]

Attribute/Model	Lasso Regression	Ridge Regression	Linear Regression
Nationality	(+) ***	(+) ***	(+) *
Position title	(-) ***	(-) ***	(-) **
Years of Experience	(+) **	(+) **	
Number of Publications	(+) *	(+) *	
Number of operations	(-) *	(-) *	(-) ***
Age Groups		(-) ***	
Gender		(-) *	
Department		(+) ***	
Patient Assessment		(+) *	
Nurses' Assessment		(-) *	
HR Assessment		(-) ***	
Supervisor Assessment		(-) *	
Number of complaints		(-) *	

Ghazzawi et al.'s study contributed to the extraction of features, as shown in Table I. All features in the models used, with 3 stars representing the most important, were evaluated with the Lasso model receiving the highest rating. Although previous studies have addressed many factors influencing physician performance, aspects still have not been thoroughly studied, such as the impact of feature selection on classification accuracy using clustering methods. This study seeks to bridge this gap by analyzing different methods to identify the most influential features. This can then cluster and make appropriate decisions for each group, contributing to improved recruitment, professional development, and administrative decision-making within healthcare organizations

A. Research Gap

1) Despite previous efforts to evaluate physician performance using traditional methods such as multi-source (MSF) assessments and questionnaires, there are clear research gaps that warrant further exploration. This study seeks to address these gaps, most notably.

2) The Lack of Use of Artificial Intelligence and Machine Learning Techniques in Physician Evaluation.

3) Most previous studies focus on traditional assessments such as patient surveys or peer reviews, without incorporating advanced techniques such as clustering to uncover hidden patterns and objectively analyze physician performance.

4) Lack of In-Depth Analysis of the Impact of Feature Selection on Clustering Accuracy.

5) Although some studies have used data analysis techniques, there is a dearth of research comparing different feature selection methods such as LASSO, regression, and dimensional analysis, and their impact on physician classification accuracy.

6) Limited Research on the Application of Clustering in the Healthcare Sector.

7) K-Means and other clustering methods have been applied in many fields, but their use in classifying physicians based on professional performance remains limited, leaving a gap in understanding how to improve evaluation quality using these techniques. Failure to Consider External Factors Influencing Ratings.

8) Some studies indicate that ratings based on patient and peer feedback may be influenced by socioeconomic factors rather than actual physician performance, calling for the development of more accurate models to mitigate bias.

9) Lack of Empirical Validation of the Impact of Ratings on Improving Healthcare Quality.

*10)*Most research is limited to data analysis without examining the actual impact of using the resulting ratings to improve management decisions and develop physician training programs.

B. Study's Contribution to Bridging the Gap

1) Applying the K-Means algorithm to classify physicians based on objective performance criteria.

2) Comparing different feature selection methods to determine the most accurate physician classification.

3) Studying the impact of various factors on the accuracy of assessments to provide a fairer and more objective model.

4) Providing recommendations for the practical application of clustering results to improve healthcare quality and administrative decision-making within medical institutions.

III. METHODOLOGY

In this study, the K-means algorithm will be applied using Python tools to cluster physician performance evaluation data, relying on three different datasets derived from the regression results in the study [6]. The study aims to analyze the effect of feature selection on the accuracy and effectiveness of the clustering process. The data was extracted from a hospital in Egypt and processed to remove outliers to ensure data quality.

A. Datasets Used

The study will rely on three sets of features: original features, 3-star-rated features, and Lasso-based features. Each set is detailed below:

1) Group 1: Original Attributes: This set contains all the original attributes collected without any dimensionality reduction. The attributes include Age, nationality, job title, years of experience, number of publications, operations, department, patient evaluation, nurse evaluation, human resources evaluation, supervisor evaluation, and complaints.

2) Group 2: Selected 3-Star Attributes: This group is based on attributes rated three-star according to their importance in the previous study analysis [6], shown in Table I. They include nationality, job title, number of operations, age groups, department, and HR assessment.

3) Group 3: Attributes Selected Using the Lasso Model: Lasso regression analysis was applied to identify the most influential attributes in the model, resulting in the selection of a smaller set of attributes: Nationality, Job Title, Years of Experience, Number of Publications, and Number of Operations.

B. Methodology for Determining the Optimal Number of Clusters (k)

1) Elbow method: This method will be used to analyze the variance within clusters and identify the inflection point that represents the balance between the number of clusters and internal consistency.

2) Silhouette index: This method will be used to assess the quality of clustering based on how distinct the clusters are from each other.

C. Finding the Optimal Number of Clusters

This study will apply unsupervised learning using the k-means algorithm to identify core clusters that may reveal common patterns in professional profiles, practice patterns, or outcomes. The features were divided into three groups, including all original features, 3-star features, and the best Lasso features, as shown in Figures [1], [2] and [3]. This division can provide valuable insights for improving resource allocation and designing effective training programs. Importantly, this approach has the potential to significantly improve the quality of healthcare, positively impacting patient outcomes. To achieve this, a systematic approach based on data-driven clustering techniques was adopted, as described in the following sections.

a) Using all original features (Optimal k=3): Fig. 1 presents the number of clusters (k=3). There is a clear distribution of the three clusters. This indicates that all original features carry strong information, allowing for the formation of three distinct clusters. This reflects a strong ability to differentiate between data, achieving high clustering accuracy. The original features are best suited when you have a diverse dataset and need good partitioning between clusters.

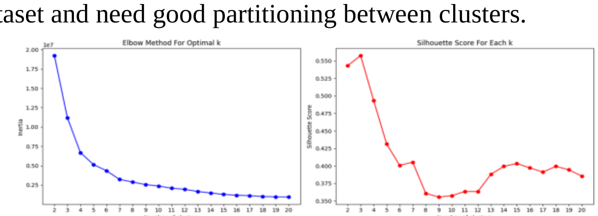

Fig. 1. Elbow Method and Silhouette Scores on original features for finding optimal k.

b) Using 3-star features (Optimal k=2): Fig. 2 shows the number of clusters (k=2). There is a less diverse distribution of clusters compared to all original features. The optimal number of clusters is only 2, indicating that the "3-star" features may lack some detail that helps differentiate between data classes. This simplification may be useful in some cases if the goal is to reduce complexity, but at the expense of some accuracy. These features may be useful if you want to simplify data or if you have a complex problem that requires less complex clustering.

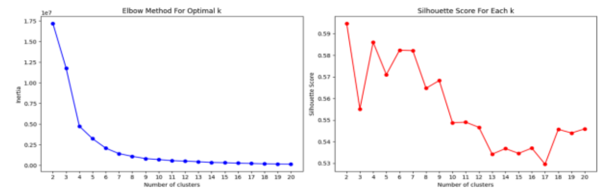

Fig. 2. Elbow Method and Silhouette Scores on 3-star features for finding optimal k.

c) Using Lasso's best features (Optimal k=3): Fig. 3 provides the number of clusters (k=3). Like the original features (k=3). The features selected by Lasso appear to select only the most important factors affecting clustering, allowing you to maintain high accuracy while reducing the number of features. It can be argued that Lasso reorders the features in a way that preserves essential details without adding additional complexity.

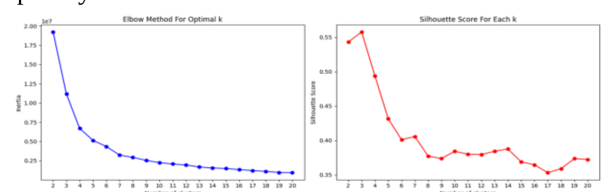

Fig. 3. Elbow Method and Silhouette Scores on Lasso's best features for finding optimal k.

The results show that both the original features and the best features selected by the Lasso model performed best at the optimal number of clusters, indicating their ability to provide accurate data segmentation. The original features provided the best cluster separation, reflecting high clustering accuracy. In contrast, the Lasso features offered an effective balance between reducing complexity and maintaining accuracy, achieving results close to those achieved using all the original features, but with fewer features.

The "three-star" features were simplified and resulted in a smaller number of clusters, which may be useful in some cases where data complexity reduction is required. However, this simplification may result in the loss of important details that may be necessary to understand subtle patterns in the data.

D. Results

The K-Means algorithm was applied to three different datasets, each with a distinct set of features. The optimal number of clusters was determined using both the Elbow method and the Silhouette index to evaluate the effectiveness of each feature set in the clustering process.

Group 1: Original Features

The results, as shown in Fig. 1, indicate that the optimal number of clusters was k=3. The Silhouette Index also recorded its highest value at this number, indicating that the original features were able to differentiate physicians into three distinct groups.

Group 2: Selected Features (Three Stars)

The results, as shown in Fig. 2 for this group, showed that k=2 was the optimal number of clusters, achieving acceptable performance. However, the Silhouette index was lower than the first group. This indicates that this group tends to oversimplify the data, which may lead to the loss of some fine detail.

Group 3: Lasso Features

The results point as Fig. 3 to k=3 as the optimal number of clusters, with a Silhouette index close to the original grouping. This indicates that the selection of these features reduced the number of variables while maintaining classification quality and accuracy.

Overall, the results indicate that feature selection has a direct impact on the number of resulting clusters and the accuracy of the clustering. Both the original and Lasso features provided accurate and meaningful clustering, while the "three-star" grouping produced a simpler model. These results demonstrate the potential for leveraging feature selection techniques to customize assessment methods more accurately and effectively in healthcare contexts. Based on these results, the study recommends using the original features due to their ability to achieve the highest clustering accuracy and distinguish clusters more clearly, making them the ideal choice when more detailed data analysis is required.

E. Analysis of Clustering Results and Appropriate Decisions

Based on The original features were adopted due to their demonstrated remarkable effectiveness in enhancing the accuracy of the clustering process and their ability to more clearly demonstrate the variation between groups, making them the ideal choice for analyzing accurate data and making personalized decisions at the level of each physician. Based on statistical analysis, it was possible to identify distinctive characteristics for each group of physicians, enabling the formulation of administrative and development decisions tailored to the nature of each group's performance.

1) Cluster 1 – Young and Mid-Earned Physicians.

Distinctive Characteristics:

Average age: 49.7 years (youngest group).

Average years of experience: 24.7 years.

Average ratings from patients, nurses, and management: High (between 4.44 and 4.84).

Average number of publications: 78.5 papers.

Medicinal error rate: 0.041 (relatively low).

Average number of operations: 800 operations.

Appropriate Decisions:

- Invest in their professional development by providing advanced training courses and mentoring programs from more experienced physicians.
- Encourage scientific research by providing grants and support for the publication of their research, as they have a good publication rate, but lower than the second group.
- Increase their administrative responsibilities, as they have good patient and nurse reviews, potentially qualifying them for future leadership roles.
- Motivate them financially and administratively by offering incentives for excellent performance, as they could be the future of the medical institution.

2) Cluster 2 – The most experienced and most surgically active physicians.

Distinctive characteristics:

Average age: 62.1 years.

Average years of experience: 37.1 years (the most experienced group).

Average ratings are very high (between 4.44 and 4.85).

Average number of publications: 136.7 papers (the highest among the three groups).

Medicinal error rate: 0.040 (the lowest among the groups).

Average number of operations: 889 operations (the highest among the groups).

Appropriate decisions:

- Keep them in leadership and supervisory roles given their extensive experience and high ratings.
- Gradually reduce the workload on them and invest their expertise in training younger physicians.
- Encourage them to focus on scientific research and participate in medical conferences to enhance the hospital's standing.
- Developing their incentive programs, such as granting job benefits to highly experienced physicians to increase their loyalty to the organization.

3) Cluster 3 – Physicians with the Least Research Involvement.

Distinctive Characteristics:

Average Age: 60.7 years.

Average Years of Experience: 35.7 years.

Average Ratings High (between 4.51 and 4.86).

Average Number of Publications: 39.6 (lowest among the three groups).

Mean Medical Error Rate: 0.047 (higher than the other two groups).

Average Number of Operations: 807.

Appropriate Decisions:

- Increase their participation in scientific research, as their publication rate is lower than that of Cluster 2. This can be achieved by providing research support or imposing research requirements for senior positions.
- Improve their medical skills and reduce errors through specialized training programs.
- Increase their involvement in academic and professional activities such as workshops and medical conferences to enhance their research experience.
- Directing them to supervise new physicians instead of focusing only on surgical procedures, to benefit from their extensive experience in training and guidance.

Finally, this analysis helps guide recruitment, training, and professional development strategies for each group of physicians more accurately. Table II provides a summary of the proposed decisions and appropriate actions for each group of physicians.

TABLE II. A SUMMARY OF THE RECOMMENDATIONS FOR EACH CLUSTER

Cluster	Main Characteristics	Proposed Decisions
Cluster 1 (young, intermediate-level physicians)	Younger age, good ratings, moderate number of research assignments, good number of operations	Professional training and development, encouragement of scientific research, management roles
Cluster 2 (more experienced and most surgically active physicians)	Highest experience, highest number of operations, highest number of research, fewest errors	Supervisory roles, gradual reduction of operations, promoting scientific research, financial incentives
Cluster 3 (physicians least involved in scientific research)	Good experience, lowest amount of research, relatively high number of errors	Professional research encouragement, training to reduce errors, integration into supervision and teaching

4) Comparison of clustering performance and results: Comparing the results, the second cluster (Cluster 2) is the most experienced and most engaged in scientific research, making it ideal for leadership and mentoring roles. The first cluster (Cluster 1) is characterized by physicians in early or mid-career, making it ideal for investing in training and professional development. The third cluster (Cluster 3) includes physicians with extensive experience but less research activity, indicating a need to enhance their involvement in academic research and improve their skills.

IV. CONCLUSION

This study indicated that selecting appropriate features directly affects the number of clusters resulting from clustering. The results showed that all original features and the best Lasso features resulted in an optimal number of clusters with a value of k=3, indicating that these features retained the essential information needed to distinguish between different physician classes. On the other hand, using simplified features reduced the number of clusters to k=2, which may be useful in some cases, but may result in the loss of some important details. In summary, this study highlights the importance of choosing appropriate features when applying clustering techniques and provides a framework that can be used in future research to analyze staff performance in medical and other fields. Furthermore, the results showed that the original features provide a clear distribution of physicians based on experience, age, and number of operations, facilitating personalized recommendations for each group. The results of this analysis can be used to support hospital decision-making in terms of developing training programs, assigning tasks, and identifying research opportunities for physicians based on their current performance.

Based on these results, the hospital recommends using cluster analysis to improve physician management and develop motivation and training programs tailored to each group, ensuring maximum utilization of available human resources.

V. STUDY LIMITATIONS

This study focuses on analyzing the impact of feature selection on the accuracy of physician clustering according to different performance criteria, using the K-Means clustering algorithm. However, some limitations should be taken into account:

1) Data limitations: The study relies on data from multiple sources, including patient reviews, peer evaluations, and administrative evaluations. This data may not be comprehensive of all aspects of physicians' professional performance, and its accuracy depends on the objectivity of the evaluators.

2) Sample scope: The data were collected from a single hospital in Egypt, which may affect the generalizability of the results to other medical institutions with different evaluation systems or work environments.

3) Influence of external factors: External factors, such as the socioeconomic environment or the nature of the health system, may affect the clustering results. These factors were not directly considered in this study.

4) Methodology used: The study relies on the K-Means clustering algorithm, which requires pre-determining the number of clusters. This may affect the accuracy of the results if the optimal number of clusters is not carefully selected.

5) Lack of validation of practical impact: The study is limited to analyzing data and testing the accuracy of the resulting classifications, without empirically validating the impact of these classifications on improving healthcare quality or actual physician performance.

VI. FUTURE WORK

Further Analysis of Selected Features: Additional studies could be conducted to analyze how the selection of different features affects clustering performance, especially in fields other than the healthcare sector.

Testing advanced clustering techniques: The use of other methods, such as hierarchical clustering or deep learning algorithms, could be studied to improve clustering results.

Scaling up the study: The study could be expanded to include other hospitals and different medical communities to compare the results and determine their generalizability.

Analyzing the impact of behavioral factors: Additional data, such as physician satisfaction and burnout levels, could be incorporated, along with their impact on the performance of different groups.

Combining cluster analysis with classification techniques: A model combining clustering and predictive classification could be developed to improve the accuracy of physician management recommendations.

This study represents a first step toward improving human resource management in hospitals using modern data analysis techniques. Further research is recommended to improve these models and enhance the accuracy of management recommendations in the future.

REFERENCES

[1] N. Brennan, M. Bryce, M. Pearson, G. Wong, C. Cooper, and J. Archer, “Understanding how appraisal of doctors produces its effects: A realist review protocol,” BMJ Open, vol. 4, no. 6, p. e005466, 2014. doi: 10.1136/bmjopen-2014-005466.

[2] J. F. Kuemmerle, “ABIM maintenance of certification 2014: Navigating the challenges to find opportunities for success,” Gastroenterology, vol. 147, no. 2, pp. 260–263, 2014. doi: 10.1053/j.gastro.2014.06.008.

[3] J. Zhang, V. L. Patel, T. R. Johnson, and E. H. Shortliffe, “A cognitive taxonomy of medical errors,” Journal of Biomedical Informatics, vol. 37, no. 3, pp. 193–204, 2004. doi: 10.1016/j.jbi.2004.04.004.

[4] Y. Shi, P. Li, X. Yu, H. Wang, and L. Niu, “Evaluating doctor performance: Ordinal regression-based approach,” Journal of Medical Internet Research, vol. 20, no. 7, p. e240, 2018. doi: 10.2196/jmir.8826.

[5] R. Xu and D. Wunsch, Clustering. John Wiley & Sons, 2008.

[6] A. M. Ghazzawi, A. O. Almagrabi, and H. M. Namankani, “Identifying influential factors behind physician performance: A machine learning approach to support decision-making,” Proceeding on Engineering Sciences (PES), in press.

[7] J. Campbell, S. Richards, A. Dickens, M. Greco, A. Narayanan, and S. Brearley, “Assessing the professional performance of UK doctors: An evaluation of the utility of the general medical council patient and colleague questionnaires,” BMJ Quality & Safety, vol. 17, no. 3, pp. 187–193, 2008. doi: 10.1136/qshc.2006.021816.

[8] S. Mirfat, M. Azzuhri, and L. Hakim, “Physician performance analysis based on individual, psychological and organizational factors in medical resume filling,” Management (JAM), vol. 16, no. 3, 2018.

[9] C. K. Cassel and S. H. Jain, “Assessing individual physician performance: Does measurement suppress motivation?” JAMA, vol. 307, no. 24, pp. 2595–2596, 2012. doi: 10.1001/jama.2012.5820.

[10] P. A. Cola and Y. Wang, “Discovering factors that influence physician scientist success in academic medical centers,” Academy of Management Proceedings, vol. 2017, no. 1, p. 12714, 2017. doi: 10.5465/ambpp.2017.12714abstract.

[11] H. Jin, J. Zhou, J. Zhang, and Y. Fu, “Factors influencing healthcare workers’ performance before and after the coronavirus disease 2019 pandemic: A bibliometric analysis with supplementary comparative analysis,” Work, Preprint, pp. 1–20, 2024. doi: 10.3233/WOR-213580.

[12] W. William, K. Kholil, T. Sukwika, and N. Ariyani, “Analysis of factors affecting the performance of medical lecturers: Case study at private ‘X’ University Indonesia,” Dinasti International Journal of Management Science, vol. 3, no. 6, pp. 1130–1145, 2022. doi: 10.31933/dijms.v3i6.1248.

[13] K. Overeem, H. C. Wollersheim, O. A. Arah, J. K. Cruijsberg, R. P. Grol, and K. M. Lombarts, “Evaluation of physicians’ professional performance: An iterative development and validation study of multisource feedback instruments,” BMC Health Services Research, vol. 12, no. 1, pp. 1–11, 2012. doi: 10.1186/1472-6963-12-80.

[14] E. Bindels, B. Boerebach, R. Scheepers, A. Nooteboom, A. Scherpbier, S. Heeneman, and K. Lombarts, “Designing a system for performance appraisal: Balancing physicians’ accountability and professional development,” BMC Health Services Research, vol. 21, no. 1, pp. 1–12, 2021. doi: 10.1186/s12913-020-06017-5.

[15] T. K. Ho and D. M. Baker, “Appraisal and revalidation,” Surgery (Oxford), vol. 30, no. 9, pp. 447–454, 2012. doi: 10.1016/j.mpsur.2012.07.004.

[16] R. D. Dias, A. Gupta, and S. J. Yule, “Using machine learning to assess physician competence: A systematic review,” Academic Medicine, vol. 94, no. 3, pp. 427–439, 2019. doi: 10.1097/ACM.0000000000002540.