Minimum Data Set (MDS) in Saudi Arabia's Medical Claims Insurance

Bashaier Al-Hamad[1], Noor Al-Dossary[1], Hayfa Al-Muhaisen[1], Mohammed Gollapalli[2], Atta Rahman[3,*], Mohammed Salih Ahmed[4]

[1]Department of Computer Information Systems, College of Computer Science and Information Technology,
Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia
[2]Department of Information Technology & Engineering, Sydney Met, Sydney, NSW 2000, Australia.
[3]Department of Computer Science, [4]Department of Computer Engineering,
College of Computer Science and Information Technology,
Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia
[*]Correspondence: aaurrahman@iau.eud.sa

Abstract
The denial of medical claims presents a critical challenge within the healthcare sector, as insurance providers decline to cover the expenses associated with providing medical assistance to policyholders in hospitals or other healthcare facilities. Claim denials often arise due to missing or inaccurate information, serving as a justification for the refusal of coverage. This research aims to explore the potential of technology in reducing the likelihood of claim denials by enabling the creation of accurate data. By implementing standardized procedures for data gathering and validation, organizations can ensure the collection and recording of necessary data, thereby enhancing data quality, reliability, customer satisfaction, operational effectiveness, and facilitating risk assessment and pricing strategies. The research community has put a lot of effort into this pertinent area, but more work is still required. There are many approaches to increase the accuracy and quality of data used in healthcare. One of these technologies is called Minimum Data Set (MDS), and it acts as a thorough database for health insurance. This method was employed in Saudi Arabia, aligning with the broader objectives outlined in Vision 2030. It gathers and assesses data related to health insurance claims, healthcare providers, beneficiaries, and the services they receive. This research study makes a valuable contribution to the existing body of knowledge by utilizing advanced data manipulation techniques on a unique dataset that has not been previously utilized for research purposes. The primary objective of this research is to maximize the potential of the MDS dataset through data manipulation. The dataset was cleaned and prepared using R Studio. A retrospective analysis was conducted to identify trends, evaluate performance, and provide insights into future decision-making. Managing research outcomes in the early stages can provide valuable insights into the context of historical data, enabling insurance companies to make more informed decisions about the future. Moreover, the researchers in this field can take this as a pilot study to lever their research journey.

Keywords:
Claim Denials, Healthcare Sector, Saudi Arabia, MDS, Insurance Company, Saudi Arabia, data wrangling.

1. Introduction

Medical claim denials pose significant challenges to healthcare reimbursements, making them a considerable burden for a doctor's practice. They can have a detrimental effect on the billing division's profitability and effectiveness. Insurance companies cannot process rejected medical claims due to data requirements that were not met, such as failing to meet specific criteria, missing data, including even one field left blank, missing modifiers, incorrect plan codes, or incorrect or missing social security numbers. As a result, these claims are never entered into their computer systems [1].

It emphasized the challenge of missing data when using health insurance claims data for advanced data analysis. The medical details in a specific claim may be limited, requiring extrapolation from later data if important outcomes are missing. For example, if an X-ray result is not recorded, later claims information might reveal that a cast was applied, indicating a fracture. Missing data can result from entry errors, code combinations, or issues with claims aggregation. Researchers need to understand the characteristics and limitations of the dataset to inform modeling decisions [1].

Another challenge highlighted the amount of medical data increases, the time required to review electronic medical records (EMR) also increases. Limited time for data review can make it difficult for doctors to provide accurate and timely treatment. It is important to explore techniques for analyzing large medical datasets [2].

Moreover, presented an example of such a technique, using nursing minimum data sets (NMDSs) to gather uniform and standardized patient data in

Manuscript received February 5, 2025
Manuscript revised February 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.2.5

healthcare services. This approach can improve healthcare quality, support resource allocation, and provide evidence of nursing work. NMDSs are particularly beneficial in electronic health records (EHRs) as they ensure uniform data collection and facilitate functions like clinical decision support and health information exchange [3].

This research paper presents a literature review of relevant studies. The study employs the Minimum Data Set (MDS) methodology as part of Saudi Arabia's vision to establish key fields in the insurance data of the Kingdom. The primary objective is to conduct investigations and analyses that fulfill the necessary requirements, ultimately reducing claim denials and improving data quality before submitting claims to the payer. The data wrangling procedures were performed using the R Studio method [4], which provides advanced statistical tools for text analysis, data cleaning, and visualization in a healthcare dataset that has not been utilized in previous research. A retrospective analysis was conducted to identify trends, evaluate performance, and offer insights into future decision-making. The main contribution of this research proposes the utilization of MDS to achieve the highest levels of accuracy, data quality, and reliability based on the findings obtained from the dataset.

This paper is organized as follows: Section 2 contains the literature review of related work, and Section 3 outlines the Method & Methodology adopted to conduct the research. Section 4 provides results and discussion, and a conclusion is offered in Section 5.

2. Review of related literature

This literature review aims to provide a systematic and comprehensive analysis of current research on MDS in insurance claim.

The authors in [5] proposed a study with the aim of identifying and analyzing Minimum Data Sets (MDSs) and comparing them to World Health Organization (WHO) guidelines. The study conducted a systematic review of the MDSs used for rare diseases in healthcare networks worldwide, including 20 studies. The results indicated a lack of terminological standardization and consistency in the structure of the MDSs. The study's significant contribution was the development of a fundamental global MDS for rare disease patient records in healthcare networks to address these issues.

In another study [6] focused on creating a Nutrition Minimum Data Set (NMDS) to standardize documentation of nutritional care in primary healthcare. The study used a two-phase methodological approach, including a systematic scoping review and inductive content analysis, to identify 32 specific items and categorize them into five distinct categories within the NMDS. These categories included physiological measurements, eating ability, dietary intake, stress-related factors, and factors indirectly influencing nutritional needs. The implementation of the NMDS for documentation could improve the quality of record-keeping, ensure continuity of care, and potentially lead to better health outcomes [5].

In [7], conducted a study that focused on creating a Minimum Data Set (MDS) for an Iranian Rheumatoid Arthritis (RA) registry. The MDS comprised a standardized collection of data elements necessary for acquiring high-quality data and establishing an integrated health information system. The study used qualitative interviews and a Delphi process to identify and validate the data elements for inclusion in the MDS. The final MDS consisted of 22 specific data elements, divided into two groups management data (such as demographic information, admission, and discharge details) and clinical data (covering patient examinations, treatment plans, and medications prescribed by physicians). This MDS provided a foundational framework for the development of an RA registry system in Iran.

In addition, [8] highlighted that the shortage and distribution disparities among behavioral health providers, the increased demand stemming from legislative changes, the mismatch between supply and demand, and the various challenges encountered in workforce planning. The proposed Minimum Data Set (MDS) is introduced as a tool to help address these issues by standardizing data collection for the behavioral health workforce.

Moreover, [9] conducted at Ahvaz University of Medical Sciences in Iran aimed to design a Minimum Data Set (MDS) for the COVID-19 registry system, recognizing the critical importance of systematic data collection during the pandemic. The research involved five phases, including assessing information requirements, identifying data elements, selecting the MDS, implementing it in a pilot COVID-19 registry system, and making necessary corrections based on the implementation experience. The resulting MDS comprises eight top groups covering various aspects of COVID-19 data, such as administrative details, disease exposure, medical history, clinical diagnostic tests, treatment outcomes, diagnosis, follow-up, and vaccination data. The study underscores that having a standardized and comprehensive MDS can facilitate the creation of a national data dictionary for COVID-19 and enhance the quality of collected data during the pandemic [10-14].

Aldhafferi et al. [15] conducted a study based on rule mining in the patient's claim data. The purpose was to extract the hidden relationship among the data tuples and potential reasons behind the rejection of claims to make informed decisions for the medical insurance companies assisting clinicians and practices. The study investigates health level 7 (HL7) related data in this regard to extract useful information particularly about the healthcare sector stakeholders including labs, practices, insurance companies and related areas [16-25].

Overall, this literature review presents key insights and outcomes in the field of MDS, contributing to the current state of knowledge and informing clinical practice.

3. Materials and Methods

R is a programming language and software environment for statistical computing and graphics, which is supported by the R Foundation for Statistical Computing. It is widely used by data miners and statisticians for developing models and analyzing data. RStudio is an integrated development environment (IDE) for R that includes a console, syntax highlighting editor with direct code execution, as well as tools for plotting, history, debugging, and workspace management [26-30].

The process starts with importing the insurance claim dataset into R Studio and exploring its structure. Missing data is handled through techniques like imputation or deletion. Inconsistencies and errors are addressed using data cleaning techniques such as string manipulation and filtering. Variable transformation may be required to convert data types or derive new variables. If multiple datasets are involved, integration is done using join functions. Validation is crucial to ensure data consistency, accuracy, and integrity, achieved through checks and exploratory analysis. Finally, the cleaned and validated dataset can be exported for further analysis or sharing with collaborators. This methodology ensures effective preparation and cleaning of research data using R Studio's capabilities [31-40].

3.1 Clinical data

The study utilized the Hypothesis healthcare dataset, which consists of extensive information on medical claims insurance for patients in Saudi Arabia. The dataset includes 12 attributes and covers 1318 records from the year 2022. Out of these records, 242 patients received emergency services while 1076 received outpatient services [41-45]. Table 1 provides a detailed explanation of the database attributes that were utilized in this research.

Table 1: Feature description

Variables	Description
Service	All laboratories, consultation, radiology, or treatment services are provided in an inpatient, outpatient setting or primary care clinic. Laboratory = 1, Consultant = 2, Radiology = 3, Service = 4.
Service_ Initial _Type	An indicator of whether the service event is the first service event for a new referral or a follow up service event. Initial event = 1, Subsequent event = 2
Emergency_ Arrival_Code	The mode of transport by which the person arrives at the emergency department, as represented by a code.
Admission_ Type	A code which identifies how the patient was admitted to hospital. Appointment = 0, ER Admission =1, Normal Admission = 2
Surgery_Flag	A qualifier to indicate whether the primary procedure code was surgical or not. None= 0, Yes = 1, No = 2

Rejected_Amount_Status	The amount where the service or claim is not accepted by insurance company. Yes = 1, No = 2
Denial_Status	Description of the rejection of the services or to the claim. Yes = 1, No = 2
Insurance_Name	Insurance company name.
Emergency_Arrival_Level	88= Level 1, 99= Level 2.
Emergency_Waiting_Time	The time elapsed in minutes for each patient from presentation in the emergency department to a service occurrence of a specified event related to service delivery.
Emergency_Waiting_Status	The status of waiting whether it's normal time or overtime.
Billed_Amount	The gross amount billed by provider for the service.

3.2 Statistical analysis

Statistical analysis plays a crucial role in comprehending and visualizing patterns within data, thereby improving the data pre-processing and modeling procedures [46-50].
Table 2 (a&b) presents the statistical analysis of the numerical attributes in the medical claim's insurance dataset. It includes details on missing values, measures of central tendency, standard deviation, minimum and maximum values, as well as the first and third quartiles.
Furthermore, Table 3 provides a concise statistical summary of the nominal features. Since the statistical analysis is not applicable on that type of attributes in the dataset.

As shown in Table 2 (a&b), there is a significant difference between the minimum values and the first quartile of billed amount attribute which demonstrates the presence of potential outliers.
Furthermore, the considerable discrepancy between the maximum values and the third quartile of the billed amount indicates the presence of an outlier. However, these outliers were treated wisely since they belong to a valid insurance company in the minimum dataset. Boxplots were constructed to further visualize the presence of existing outliers and other characteristics of the dataset, as illustrated in Figure 1.

Table 2 (a): Numerical attributes statistical description

Feature	count	mean	std	min
Service	1318	3.83	0.629	1
Service_Initial_Code	1318	1.82	0.385	1
Emergency_Arrival_Code	1318	90.0	4.26	88
Admission_Type	1318	0.192	0.546	0
Billed_Amount	1318	964.	6433.	2
Surgery_Flag	1318	0.149	0.495	0
Rejected_Amount_Status	1318	1.90	0.298	1
Denial_Status	1318	1.93	0.257	1
Emergency_Waiting_Time	1318	5.04	14.5	1

Table 2 (b): Numerical attributes statistical description

Feature	25%	50%	75%	max
Service	4	4	4	4
Service_Initial_Code	2	2	2	2
Emergency_Arrival_Code	88	88	88	99
Admission_Type	0	0	0	2
Billed_Amount	168.	300	630	222283
Surgery_Flag	0	0	0	2
Rejected_Amount_Status	2	2	2	2
Denial_Status	2	2	2	2
Emergency_Waiting_Time	0	0	0	128

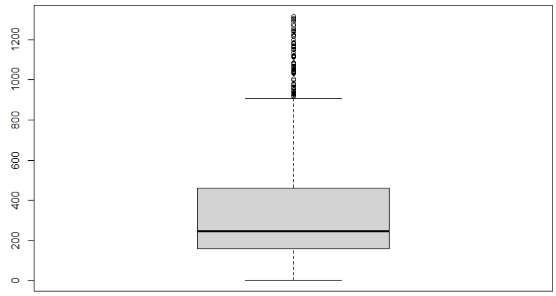

Figure 1: Boxplot for Billed Amount

Likewise, Table 3 provides the description of nominal (non-numerical) attributes. It is worth noting that the total count of each attribute related value is 1318 each.

Table 3: Nominal Attributes Statistical Analysis

Feature	Values
Insurance_Name	Laboratory (44), Consultant (34), Radiology (29), Service (1211).
Emergency_Arrival_Level	Level 1(242), Level 2 (1076)
Emergency_Waiting_Status	Overtime (90), Normal Time (152), No Waiting Time (1076).

4. Retrospective analysis

This section comprehensively details the retrospective study conducted as a first step on each of the target questions for Claim Health Insurance data through descriptive statistics. The study aimed to understand the determinant factors associated with Insurance companies to provide information about the services offered by different medical insurance companies and it impacts on different patients and hospital [51-60].

4.1 Count Services by Medical Insurance Company

The diverse services provided by different insurance companies in the kingdom can be a valuable source of information for identifying market gaps and opportunities for innovation and differentiation. To gain a better understanding, the analysis focuses on the distribution of services among various medical insurance companies. By categorizing services into different categories, such as laboratory, consultants, radiology, and normal service, insights can be obtained into how these services are distributed among insurance companies like Bupa, Tawuniya, Sabic, and other [61-65].

This analysis is presented in the form of a bar plot in Figure 2, which shows the higher or lower service counts in each category. First, consider BUPA with a count of 25.83% for normal services, followed by Tawuniya and others with the lowest count at 0.09% for radiology services.
We further figured out that SABIC has the highest number of normal services at 46.31% compared to others. This suggests that by aligning insurance companies' strategies with the evolving services needs and demands of patients, it can enhance patients' satisfaction and drive growth.

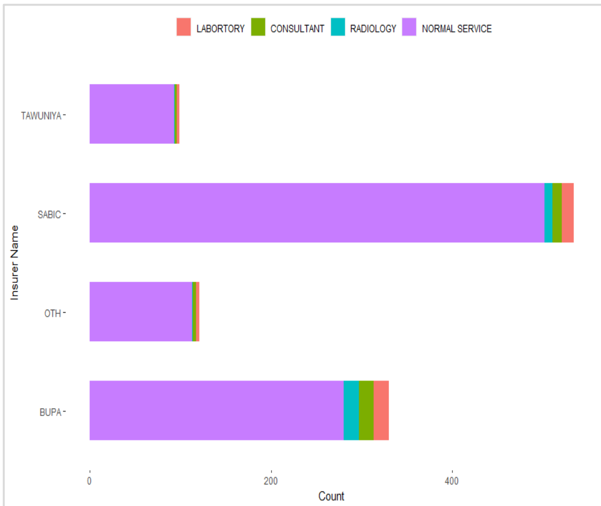

Figure 2: Count Services by Medical Insurance Company

4.2 The Percentage of Admission Type

Understanding the varying demands for different types of admissions is crucial for insurance companies to effectively allocate resources, streamline processes, and reduce waiting times [66-70]. In this retrospective study, the percentage of different admission types at a hospital was analyzed using the data set demonstrated via pie chart.

Figure 3 illustrates the division of admission types into three sections. The analysis revealed that ER admission accounts for 4%, followed by normal admission at 3%. Interestingly, the "Appointment" category had the highest percentage among all admission types, with a staggering 93%. This highlights the importance of optimizing appointment booking systems and improving scheduling processes to meet the high demand. These findings have significant implications for insurance companies as they provide valuable insights into the demand for various admission types, enabling them to optimize operations and enhance the patient's experience.

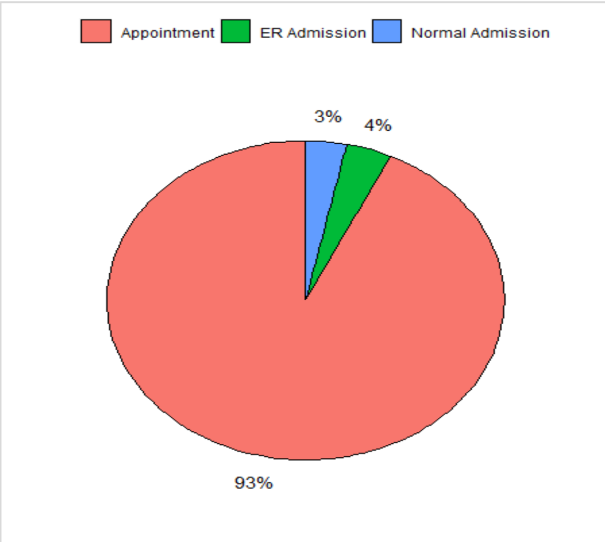

Figure 3. The Percentage of Admission Type

4.3 Amount Billed by Insurance Company

Analyzing the billed amounts for different services and insurance companies is crucial for identifying trends, evaluating service profitability, and making informed decisions about resource allocation and pricing strategies. Figure 4 shows the boxplot data, indicating that Bupa has a higher median billed amount compared to TAWUNIYA and others, suggesting a potential trend. Additionally, SABIC stands out with the highest median billed amount among other insurance companies. That shows the relative rate as well as the responsiveness level.

Table 4 summarizes descriptive statistics for the billed amount by service after removing outliers. It includes information on the number of observations, mean, standard deviation, median, trimmed mean, minimum and maximum values, range, skewness, kurtosis, and standard error for each service category. These statistics offer insights into the distribution, variability, and central tendency of the billed amounts across various insurance companies.

The cost of the four services varies significantly, with radiology being the most expensive. The distribution of the billed amounts is right-skewed, indicating a few outliers that drive up the mean. The trimmed mean and max provide relatively more robust measures of central tendency and dispersion, respectively.

Skewness and kurtosis describe the shape of the distribution. These findings provide valuable insights for insurance companies to understand revenue generation from different service categories and identify areas for improvement.

The analysis further helps in resource allocation by identifying high-revenue services that require relatively more attention. It also aids in pricing strategies by considering billed amounts for different services, enabling companies to set competitive prices while ensuring their respective profitability. Further, the companies can associate the prices with other factors to gain more attention. For instance, age group, gender and type of salaried class or businessmen and self-employed patients.

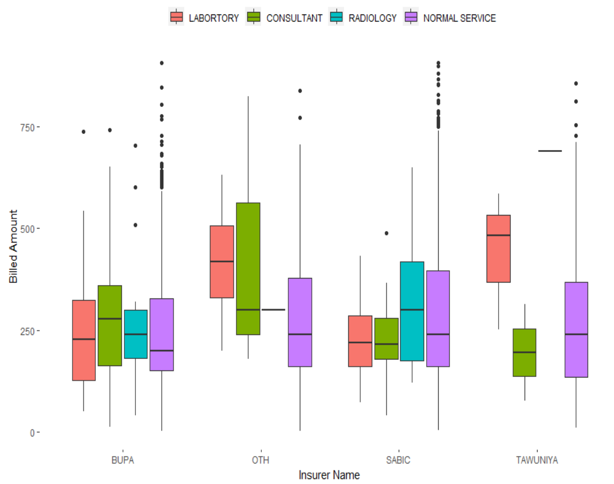

Figure 4. Amount Billed by Insurance Company

Table 4. Descriptive Statistics for Billed Amount by Service

Service	Lab	Consult	Radiology	Normal Service
Vars	1	1	1	1
n	37	31	28	988
Mean	280.95	286.26	310.00	287.19
Sd	171.02	195.07	186.49	199.83
Median	246.0	228.0	240.0	233.5
Trimmed	266.94	262.04	296.92	264.86
Mad	152.71	126.02	106.75	144.55
Min	50	13	40	2
Max	738	824	704	907
Range	688	811	664	905
Skew	0.78	1.11	0.89	0.97
Kurtosis	-0.13	0.75	-0.49	0.16
Se	28.11	35.04	35.24	6.36

4.4 Rejection rate

The insurance company must make well-informed decisions regarding insurance partnerships, negotiate favorable terms, and develop strategies to minimize claim rejections. In this study, Figure 5 demonstrates an overview of the distribution of the total amount among different insurers can be seen. Bupa accounts for 30.4% of the total, while Tawaniya represents 11.2%. Additionally, other insurers make up 9.13% of the total. Notably, Sabic holds the largest portion with 49.3% of the total claim rejections. This indicates that the total rejection amount in the year 2022 accounts for 9.87% of the total claims made as illustrated in Figure 6.

These findings highlight the variations in performance among insurance companies in terms of claim processing, with Sabic experiencing the highest amount of rejected billed amounts. This information can lead to potential time and cost savings, improved customer satisfaction, and a better understanding of insurer performance. Insurance companies can utilize this knowledge to enhance their insurance claims processing systems and optimize their insurance relationships.

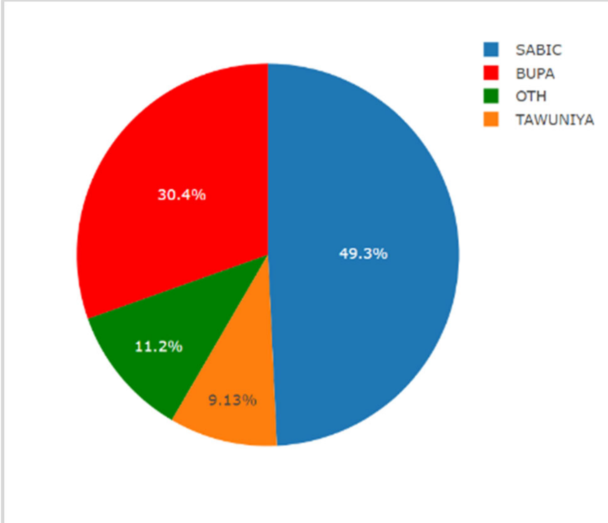

Figure 5. Rejection rate by each insurance company

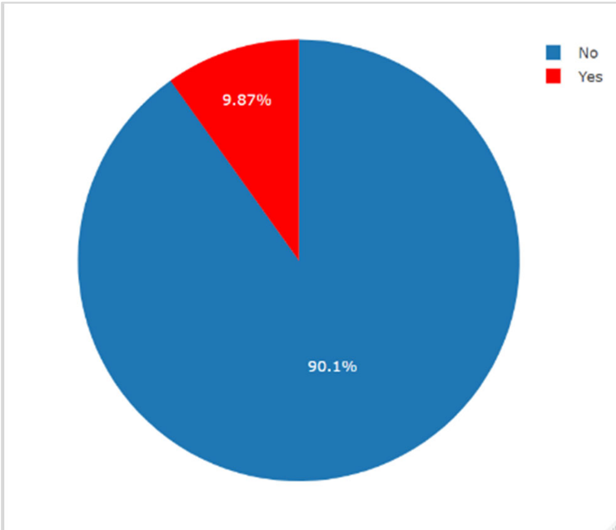

Figure 6. Overall rejection rate by all companies

4.5 Denial Status by Services and Billed Amount

Insurance companies must identify areas for improvement, reduce claim denials, and optimize revenue management. Figure 7 illustrates the denial status of services and the corresponding billed amounts. The box plot shows that consultant services have the highest median denial status, followed by normal services. The goal of this analysis is to provide insurance companies with valuable insights to make informed decisions and implement strategies to minimize claim denials. By understanding the denial status and billed amounts, insurance companies can focus on areas with higher denial rates, enhance operational efficiency, and mitigate financial losses. These insights offer potential cost savings through reduced claim denials, improved revenue management by addressing denial patterns, and increased operational efficiency. They empower insurance companies to make data-driven decisions and optimize their operations for better financial performance.

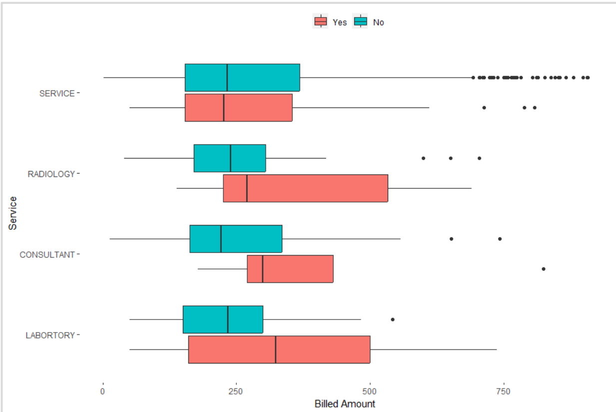

Figure 7. Denial by Services and Billed Amount

5. Discussion

Medical claims denial has emerged as a significant challenge in the healthcare industry. Issues such as inaccurate or incomplete documentation, coding errors, lack of standardized procedures, and limited communication between payers and providers contribute to claim denials. These challenges result in financial losses for healthcare providers and hinder timely access to care for patients [71-75].

To address these challenges, this study introduces the Minimum Data Set (MDS) methodology. The MDS methodology involves gathering and analyzing standardized data related to health insurance claims, healthcare providers, beneficiaries, and services received. Implementing the MDS methodology can enhance data quality, reliability, and accuracy, leading to a reduction in claim denials. This study discusses the potential benefits of implementing the MDS methodology and emphasizes the importance of standardized data gathering and validation procedures to improve data quality and reliability. By ensuring accurate and complete documentation, healthcare organizations can enhance customer satisfaction, operational effectiveness, and financial performance. The MDS methodology plays a crucial role in reducing claim denials by providing a standardized framework for data collection and analysis. It ensures consistent capture of all relevant information, minimizing the chances of incomplete or inaccurate documentation. With standardized procedures in place, healthcare organizations can streamline their claim submission processes and minimize coding errors. In this study, retrospective analysis conducted using R Studio enables a comprehensive examination of healthcare datasets. It allows researchers to identify patterns or trends that may contribute to claim denials. For instance, data analysis can reveal specific coding errors consistently leading to claim denials, providing insights for targeted interventions or training programs. Additionally, retrospective analysis helps evaluate the performance of healthcare organizations in terms of claim denials by analyzing key performance indicators such as denial rates or average processing times. This information enables informed decision-making regarding process changes or resource allocation [76-80]. Further, the proposed methodology highlights the significance of standardized data gathering, validation procedures, and data analysis in improving data quality, reliability, and operational effectiveness in healthcare organizations. The findings contribute to existing knowledge and provide valuable insights for future research and decision-making in this field [81-85].

6. Conclusion and recommendations

Claim denials have a significant impact on the healthcare sector in Saudi Arabia, resulting from issues such as inaccurate or incomplete documentation, coding errors, lack of standardized procedures, and limited communication between payers and providers. Researchers have developed methods to address these challenges, which have significant implications for healthcare organizations, leading to financial losses and potential harm to patient care. This research emphasizes the importance of implementing the Minimum Data Set (MDS) methodology and highlights the need for standardized data gathering and validation procedures to improve data quality and reliability. A statistical retrospective analysis was conducted on a unique healthcare dataset to identify trends, evaluate performance, and provide insights for future decision-making. The findings contribute to the existing knowledge on medical claim denials and demonstrate the potential of data analysis in reducing claim denials. As part of the recommendations for future work, further research is needed to explore the long-term effectiveness of the MDS methodology in reducing claim denials and its impact on patient outcomes. Also, future studies could investigate other technological solutions and strategies that can complement the MDS methodology in addressing the challenges of medical claim denials. In short, this study contributes to the existing body of knowledge on medical claim denials and provides a foundation for future research in this field. It underscores the importance of addressing these challenges to enhance the efficiency and effectiveness of healthcare systems.

References

[1] R. Konrad, W. Zhang, M. Bjarndóttir and R. Proaño, "Key considerations when using health insurance claims data in advanced data analyses: an experience report," Health Systems, vol. 9, p. 317–325, 2020.

[2] M. El Khatib, S. Hamidi, I. Al Ameeri, H. Al Zaabi and R. Al Marqab, "Digital disruption and big data in healthcare-opportunities and challenges," ClinicoEconomics and Outcomes Research, p. 563–574, 2022.

[3] F. Freguia, M. Danielis, R. Moreale and A. Palese, "Nursing minimum data sets: Findings from an umbrella review," Health Informatics Journal, vol. 28 (2), 2022.

[4] S. Hussain, L. J. Muhammad and A. Yakubu, "Mining social media and DBpedia data using Gephi and R," Journal of Applied Computer Science & Mathematics, vol. 12, p. 14–20, 2018.

[5] F. A. Bernardi, B. Mello de Oliveira, D. Bettiol Yamada, M. Artifon, A. M. Schmidt, V. Machado Scheibe, D. Alves and T. M. Félix, "The Minimum Data Set for Rare Diseases: Systematic Review," Journal of Medical Internet Research, vol. 25, p. e44641, 2023.

[6] S. J. Håkonsen, P. U. Pedersen, A. Bygholm, M. D. J. Peters and M. Bjerrum, "Speaking the same language: Development of a nutrition minimum data set for healthcare professionals in primary healthcare," Health Informatics Journal, vol. 26, p. 248–263, 2020.

[7] M. Langarizadeh, N. Mehrabi, T. Azadi, E. Mehraeen and A. Ahmadzadeh, "Developing A Minimum Data Set for A Rheumatoid Arthritis Registry in Iran," Mediterranean Journal of Rheumatology, vol. 33, p. 55, 2022.

[8] A. J. Beck, P. M. Singer, J. Buche, R. W. Manderscheid and P. Buerhaus, "Improving data for behavioral health workforce planning: development of a minimum data set," American journal of preventive medicine, vol. 54, p. S192–S198, 2018.

[9] J. Zarei, M. Badavi, M. Karandish, M. Haddadzadeh Shoushtari, M. Dastoorpoor, F. Yousefi, H. Raji and M. Cheraghi, "A study to design minimum data set of COVID-19 registry system," BMC Infectious Diseases, vol. 21, p. 1–13, 2021.

[10] A. Rahman et al., "Supervised machine learning-based prediction of COVID-19," Computers, Materials and Continua, vol. 69, no. 1, pp. 21-34, 2021.

[11] R. Zagrouba et al., "Modelling and Simulation of COVID-19 Outbreak Prediction Using Supervised Machine Learning." Computers, Materials & Continua, vol. 66, no. 3, pp. 2397-2407, 2021.

[12] R. A. Naqvi et al., "Coronavirus: A "Mild" Virus Turned Deadly Infection," Computers, Materials & Continua, vol. 67, no. 2, pp. 2631-2646, 2021.

[13] A. Alqarni and A. Rahman, "Arabic Tweets-Based Sentiment Analysis to Investigate the Impact of COVID-19 in KSA: A Deep Learning Approach," Big Data and Cognitive Computing, vol. 7, no. 1, pp. 1-16, 2023.

[14] M. S. Ahmed et al., "Joint Diagnosis of Pneumonia, COVID-19, and Tuberculosis from Chest X-ray Images: A Deep Learning Approach." *Diagnostics*, vol. 13(1), p. 2562, 2023.

[15] N. Aldhafferi, A. Alqahtani, A. Rahman, and M. Azam, "Constraint Based Rule Mining in Patient Claim Data," Journal of Computational and Theoretical Nanoscience 15 (3), 164-171, 2018.

[16] A. Rahman and J. Alhiyafi, "Health Level Seven Generic Web Interface," Journal of Computational and Theoretical Nanoscience 15 (4), 1261, 2018.

[17] M. Zaheer, I.M. Qureshi, A. Rahman, J. Alhiyafi, and M.Z. Muzaffar, "Improved and Secure Differential LSB Embedding Steganography," Journal of Information Assurance and Security 11 (5), 170-178, 2017

[18] M. Al-Dossary, M.S. Alshahrani, A.F. Altamimi, M. Gollapalli, A. Rahman, "Evaluating Mental Health and Well-Being through Social Media Analysis," IJCSNS - International Journal of Computer Science and Network Security, 24 (11):21-34, 2024.

[19] A. Rahman et al., "A neuro-fuzzy approach for user behaviour classification and prediction," Journal of Cloud Computing, vol. 8, no. 1, pp. 1-15, 2019.

[20] G. Zaman, H. Mahdin, K. Hussain, A. Rahman, J. Abawajy and S. A. Mostafa, "An ontological framework for information extraction from diverse scientific sources," IEEE Access, vol. 9, pp. 42111-42124, 2021.

[21] N.A. Sajid, M. Ahmad, A. Rahman, G. Zaman, M.S. Ahmed, N. Ibrahim et al., "A Novel Metadata Based Multi-Label Document Classification Technique," Computer Systems Science and Engineering 46 (2), 2195-2214, 2023.

[22] S. Arooj, M. F. Khan, T. Shahzad, M. A. Khan, M. U. Nasir et al., "Data fusion architecture empowered with deep learning for breast cancer classification," Computers, Materials & Continua, vol. 77, no.3, pp. 2813–2831, 2023.

[23] F. Jan et al., "Assessing acetabular index angle in infants: a deep learning-based novel approach," J. Imaging, vol. 9, no. 1, p. 242, 2023, doi: 10.3390/jimaging9110242.

[24] M. I. B. Ahmed et al., "Personal protective equipment detection: a deep-learning-based sustainable approach," Sustainability, vol. 15, no. 1, p. 13990, 2023.

[25] M.S. Farooq, S. Abbas, A. Rahman, K. Sultan, M.A. Khan, A. Mosavi, "A Fused Machine Learning Approach for Intrusion Detection System," Computers, Materials & Continua 74 (2), 2607–2623, 2023.

[26] M. I. B. Ahmed et al., "A deep-learning approach to driver drowsiness detection," Safety, vol. 9, no. 65, 2023.

[27] A. Rahman et al., "Geo-Spatial Disease Clustering for Public Health Decision Making," Informatica, vol. 46, no. 6, pp. 21-32, 2022.

[28] M.A. Qureshi et al., "Aspect Level Songs Rating Based Upon Reviews in English." Computers, Materials & Continua 74 (2), 2589-2605, 2023.
[29] T. A. Khan et al., "Secure IoMT for Disease Prediction Empowered with Transfer Learning in Healthcare 5.0, the Concept and Case Study," IEEE Access, vol. 11, pp. 39418-39430, 2023, doi: 10.1109/ACCESS.2023.3266156.
[30] W. H. Hantom and A. Rahman, "Arabic Spam Tweets Classification: A Comprehensive Machine Learning Approach," AI, vol. 5, no. 3, pp. 1049-1065, 2024.
[31] A. Rahman et al., "Rainfall Prediction System Using Machine Learning Fusion for Smart Cities." Sensors, vol. 22, no. 9, p. 3504, 2022, https://doi.org/10.3390/s22093504.
[32] M. Farooqui et al., "A Deep Learning Approach to Industrial Corrosion Detection," CMC-Computers, Materials & Continua, vol. 81, no. 1, pp. 1-19, 2024.
[33] M. N. Alnuaimi et al., "Transfer Learning Empowered Skin Diseases Detection in Children," Computer Modeling in Engineering & Sciences, vol. 141, no. 3, pp. 1-15, 2024.
[34] S.M. Alotaibi, Atta-ur-Rahman, M.I. Basheer, and M.A. Khan, "Ensemble Machine Learning Based Identification of Pediatric Epilepsy," Comput. Mater. Contin., vol. 68, no. 1, pp. 149-165, 2021.
[35] A. Rahman et al., "Diabetic Retinopathy Detection: A Hybrid Intelligent Approach," Computers, Materials and Continua, vol. 80, no. 3, pp. 4561-4576, 2024.
[36] M.T. Naseem, I.M. Qureshi, A. Rahman, M.Z. Muzaffar, "Robust and Fragile Watermarking for Medical Images using Redundant Residue Number System and Chaos," Neural Network World 30 (3), 177-192, 2020.
[37] A. Rehman et al., "Modelling, simulation, and optimization of diabetes type II prediction using deep extreme learning machine," Journal of Ambient Intelligence and Smart Environments, vol. 12, no. 2, pp. 125-138, 2020.
[38] N. M. Ibrahim et al., "Transfer Learning Approach to Seed Taxonomy: A Wild Plant Case Study." Big Data Cogn. Comput., vol. 7, no. 128, 2023.
[39] A. Alhashem et al., "Diabetes Detection and Forecasting using Machine Learning Approaches: Current State-of-the-art," IJCSNS - International Journal of Computer Science and Network Security, vol. 23, no. 10, pp. 199-208, 2023.
[40] M Gul, IA Khan, G Zaman, A Rahman et al., "A Game-Theoretic Approach to Safe Crowd Evacuation in Emergencies," CMC-Computers, Materials & Continua 79 (1), 1631-1657, 2024.
[41] Alabbad, D.A.; Ajibi, S.Y.; Alotaibi, R.B.; Alsqer, N.K.; Alqahtani, R.A.; Felemban, N.M.; Rahman, A.; Aljameel, S.S.; Ahmed, M.I.B.; Youldash, M.M. Birthweight Range Prediction and Classification: A Machine Learning-Based Sustainable Approach. Mach. Learn. Knowl. Extr. 2024, 6, 770-788.
[42] I.A. Qureshi et al., "GFuCWO: A genetic fuzzy logic technique to optimize contention window of IEEE-802.15. 6 WBAN," Ain Shams Engineering Journal, 102681, 2024.
[43] D.M. Althawadi et al., "Exploring Efficient Solutions for the 0/1 Knapsack Problem," IJCSNS - International Journal of Computer Science and Network Security 24(2): 15-24, 2024.
[44] M. M. Qureshi, F. B. Yunus, J. Li, A. Ur-Rahman, T. Mahmood and Y. A. A. Ali, "Future Prospects and Challenges of On-Demand Mobility Management Solutions," in IEEE Access, vol. 11, pp. 114864-114879, 2023.
[45] I. Alrashide et al., "AIMS: AI based Mental Healthcare System," IJCSNS - International Journal of Computer Science and Network Security 23(12): 225-234, 2023.
[46] M. Gollapalli, A. -U. Rahman, A. Osama, A. Alfaify, M. Yassin and A. Alabdullah, "Data Mining and Visualization to Understand Employee Attrition and Work Performance," 2023 3rd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, 2023, pp. 149-154.
[47] Musleh DA, Olatunji SO, Almajed AA, Alghamdi AS, Alamoudi BK, Almousa FS, Aleid RA, Alamoudi SK, Jan F, Al-Mofeez KA, et al. Ensemble Learning Based Sustainable Approach to Carbonate Reservoirs Permeability Prediction. Sustainability. 2023; 15(19):14403.
[48] R.A. Qamar, M. Sarfraz, A. Rahman, S.A. Ghauri, "Multi-Criterion Multi-UAV Task Allocation under Dynamic Conditions," Journal of King Saud University-Computer and Information Sciences 35 (9), 101734, 2023.
[49] N. AlDossary, S. AlQahtani, R. Alzaher, A. Rahman, "SYN Flood DoS Detection System Using Time Dependent Finite Automata," International Journal of Computer Science & Network Security 23 (6), 147-154, 2023.
[50] Z. Alsadeq, H. Alubaidan, A. Aldweesh, A. Rahman, T. Iqbal, "A Proposed Model for Supply Chain using Blockchain Framework," IJCSNS - International Journal of Computer Science and Network Security 23(6): 91-98, 2023.
[51] A. Albassam et al., "Integration of Blockchain and Cloud Computing in Telemedicine and Healthcare," IJCSNS, 23 (6): 17-26, 2023.
[52] Sajid, N.A. et al. Single vs. Multi-Label: The Issues, Challenges and Insights of Contemporary Classification Schemes. Appl. Sci. 2023, 13, 6804.
[53] Gollapalli, M. et al. SUNFIT: A Machine Learning-Based Sustainable University Field Training Framework for Higher Education. Sustainability 2023, 15, 8057.
[54] Talha, M.; Sarfraz, M.; Rahman, A.; Ghauri, S.A.; Mohammad, R.M.; Krishnasamy, G.; Alkharraa, M. Voting-Based Deep Convolutional Neural Networks (VB-DCNNs) for M-QAM and M-PSK Signals Classification. Electronics 2023, 12, 1913.
[55] Musleh, D.; Alotaibi, M.; Alhaidari, F.; Rahman, A.; Mohammad, R.M. Intrusion Detection System Using Feature Extraction with Machine Learning Algorithms in IoT. J. Sens. Actuator Netw. 2023, 12, 29.
[56] Alghamdi, A.S.; Rahman, A. Data Mining Approach to Predict Success of Secondary School Students: A Saudi Arabian Case Study. Educ. Sci. 2023, 13, 293.
[57] Alhaidari, F., Rahman, A. & Zagrouba, R. Cloud of Things: architecture, applications and challenges. J Ambient Intell Human Comput 14, 5957–5975 (2023).
[58] Rahman, A. GRBF-NN based ambient aware realtime adaptive communication in DVB-S2. J Ambient Intell Human Comput 14, 5929–5939 (2023).
[59] Ahmad, M. et al. Enhanced query processing over semantic cache for cloud based relational databases. J Ambient Intell Human Comput 14, 5853–5871 (2023).
[60] Basheer Ahmed, M.I et al., A Real-Time Computer Vision Based Approach to Detection and Classification of Traffic Incidents. Big Data Cogn. Comput. 2023, 7, 22.
[61] M Jamal, NA Zafar, D Musleh, MA Gollapalli, S Chabani, "Modeling and Verification of Aircraft Takeoff Through Novel Quantum Nets." Computers, Materials & Continua 72 (2), 3331-3348, 2022.
[62] Q Hussain, ASM Noor, MM Qureshi, J Li, A Rahman, A Bakry et al., "Reinforcement learning based route optimization model to enhance energy efficiency in internet of vehicles," Scientific Reports 15, 3113, 2025.
[63] Youldash, M., Rahman, A., Alsayed, M., Sebiany, A., Alzayat, J. et al. (2025). Deep learning empowered diagnosis of diabetic retinopathy. Intelligent Automation & Soft Computing, vol. 40, no. 1, pp. 125–143.
[64] Youldash, M., Rahman, A., Alsayed, M., Sebiany, A., Alzayat, J., Aljishi, N., Alshammari, G., & Alqahtani, M. (2024). Early Detection and Classification of Diabetic Retinopathy: A Deep Learning Approach. AI, 5(4), 2586-2617. https://doi.org/10.3390/ai5040125.
[65] W.Q. Li, X.X. Han, Z.B. Lin, A. Rahman, "Enhanced pest and disease detection in agriculture using deep learning-enabled drones,"

Acadlore Transactions on AI and Machine Learning (ATAIML) 3 (1), 1-10, 2024.
[66] C. Li, Y. Cai, L.J. Habeeb, A. Rahman, Ritzkal, "Enhancing Fall Risk Assessment in the Elderly: A Study Utilizing Transfer Learning in an Improved EfficientNet Network with the Gramian Angular Field Technique," Healthcraft Frontiers 1 (1), 1-14, 2023.
[67] Rahman, A.-u., Alqahtani, A., Aldhafferi, N., Nasir, M. U., Khan, M. F., Khan, M. A., & Mosavi, A. (2022). Histopathologic Oral Cancer Prediction Using Oral Squamous Cell Carcinoma Biopsy Empowered with Transfer Learning. Sensors, 22(10), 3833.
[68] M Mahmud, A Rahman, M Lee, JY Choi, "Evolutionary-based image encryption using RNA codons truth table," Optics & Laser Technology 121, 105818, 2020.
[69] Ghazal, Taher M et al. "Supervised Machine Learning Empowered Multifactorial Genetic Inheritance Disorder Prediction." Computational intelligence and neuroscience vol. 2022 1051388. 31 May. 2022,
[70] Adnan Khan M, Abbas S, Atta A, Ditta A, Alquhayz H, Farhan Khan M, et al. Intelligent cloud based heart disease prediction system empowered with supervised machine learning. Comput Mater Contin. 2020;65(1):139–151.
[71] S Dash, S BISWA, D BANERJEE, A Rahman, "Edge and Fog Computing in healthcare – a review," Scalable Computing: Practice and Experience 20 (2), 191-205, 2019.
[72] Rahman, Au., Dash, S. & Luhach, A.K. Dynamic MODCOD and power allocation in DVB-S2: a hybrid intelligent approach. Telecommun Syst 76, 49–61 (2021).
[73] Rahman, A. GRBF-NN based ambient aware realtime adaptive communication in DVB-S2. J Ambient Intell Human Comput 14, 5929–5939 (2023).
[74] Ibrahim, N.M., Gabr, D.G.I., et al. A deep learning approach to intelligent fruit identification and family classification. Multimed Tools Appl 81, 27783–27798 (2022).
[75] F. Alhaidari et al., "Intelligent Software-Defined Network for Cognitive Routing Optimization Using Deep Extreme Learning Machine Approach," Comput. Mater. Contin., vol. 67, no. 1, pp. 1269–1285, 2021.
[76] F. Alhaidari et al., "ZeVigilante: Detecting Zero-Day Malware Using Machine Learning and Sandboxing Analysis Techniques," Computational Intelligence and Neuroscience 2022, 1615528, 15 pages, 2022.
[77] M. Gollapalli et al., "Appendicitis Diagnosis: Ensemble Machine Learning and Explainable Artificial Intelligence-Based Comprehensive Approach." Big Data and Cognitive Computing, 8(9), 108, 2024. Doi: 10.3390/bdcc8090108.
[78] S. Abbas, S.A. Raza, M.A. Khan, A. Rahman, K. Sultan, and A. Mosavi, "Automated File Labeling for Heterogeneous Files Organization Using Machine Learning." Computers, Materials & Continua 74 (2), 3263-3278, 2023.
[79] M. Zaheer, I.M. Qureshi, K. Sultan, A. Rahman, M.Z. Muzaffar, R. Alnanih, "High Capacity Image Steganography Based on Prime Series Representation and Payload Redundancy Removal," Journal of Information Assurance and Security 14 (2), 40-47, 2019.
[80] A. Rahman, S.A. Alrashed, A. Abraham, "User Behaviour Classification and Prediction Using Fuzzy Rule Based System and Linear Regression," Journal of Information Assurance and Security 12 (3), 86-93, 2017.
[81] D.A. Almubayedh et al., "Quantum bit commitment on IBM QX," Quantum Information Processing 19 (55), 2020.
[82] H.M. Faisal, M.A. Tariq, A. Rahman, A. Alghamdi, N. Alowain, "A Query Matching Approach for Object Relational Databases Over Semantic Cache," Application of Decision Science in Business and Management 1, 2019.
[83] K. Sultan, I.M. Qureshi, A. Rahman, B.A. Zafar, M. Zaheer, "CSI Based Multiple Relay Selection and Transmit Power Saving Scheme for Underlay CRNs Using FRBS and Swarm Intelligence," International Journal of Applied Metaheuristic Computing (IJAMC) 10 (3), 1-18, 2019.
[84] A. Rahman, "Memetic computing based numerical solution to Troesch problem," Journal of Intelligent and Fuzzy Systems 36 (6), 1-10, 2019.
[85] A. Rahman, "Optimum information embedding in digital watermarking," Journal of Intelligent & Fuzzy Systems 37 (1), 553-564, 2019.