
APPLICATION OF DATA MINING TECHNIQUES IN HEALTHCARE: IDENTIFYING INTER-DISEASE RELATIONSHIPS THROUGH ASSOCIATION RULE MINING

Haryanto¹, Hadi Winarto², Christina Juliane³

STMIK LIKMI, Jawa Barat, Indonesia

jr.haryanto@gmail.com¹, hdwinarto@gmail.com², christina.juliane@likmi.ac.id³

ABSTRACT

This research focuses on the application of data mining techniques in a healthcare environment by utilizing patient visit data from Hospital X, coded with ICD-10 diagnoses. The purpose of this study is to explore the application of data mining techniques in a healthcare environment, specifically to identify the relationship between diseases using patient visit data from X Hospital. This research utilizes the FP-Growth algorithm method followed by Association Rule Mining to find frequent occurrences of diseases in the data set. The research process involved data pre-processing, transformation into binary format, and careful parameter setting (minimum support 0.95 and confidence 0.9). The results showed a strong association between chronic conditions such as hypertension and diabetes, which are prevalent in the patient population. This association provides insight into potential comorbidities and may assist healthcare providers in improving diagnosis accuracy and treatment effectiveness. This research has implications for the application of data mining techniques, demonstrating its potential in improving predictive analytics in healthcare and strategic planning. This approach not only aids in the efficient allocation of healthcare resources, but also aligns with the broader goal of improving personalized patient care.

Keywords: Data Mining, Healthcare, FP-Growth Algorithm, Association Rule Mining, ICD-10.

Corresponding Author: Haryanto

E-mail: jr.haryanto@gmail.com



INTRODUCTION

The healthcare industry has undergone significant transformations over the past few decades, primarily driven by advancements in technology and data management (Wang et al., 2018). The digitalization of health records and the systematic collection of patient data have improved the quality of care and opened new avenues for analyzing health trends and managing diseases more effectively (Senbekov et al., 2020).

In Indonesia, a nation characterized by its vast geographic spread and diverse population, the management and analysis of health data present unique challenges and opportunities (Hilbert, 2016). The country's healthcare system must handle various diseases and health conditions, from communicable diseases common in tropical climates to non-communicable diseases arising from lifestyle changes. Tracking and analyzing these conditions accurately is crucial for effective public health management and policy-making (Groseclose & Buckeridge, 2017).

Within this context, the International Classification of Diseases, Tenth Revision (ICD-10), plays a pivotal role. As a global standard for reporting diseases and health conditions, ICD-10 provides a comprehensive system that allows healthcare professionals and policymakers to systematically store, retrieve, and analyze health information (Alyahya & Khader, 2019). This ensures consistency

and accuracy in the data collected, facilitating effective national and international health strategies, disease management, and epidemiological studies (Alonge et al., 2019).

Data mining has emerged as a powerful tool in healthcare, offering the potential to extract valuable insights from vast amounts of data (Gul et al., 2021). Some methods and algorithms can be implemented in the healthcare industry. One such method is association rule mining, which aims to discover meaningful relationships between diseases or health conditions (Behzadnia et al., 2020). Healthcare providers can uncover hidden patterns and relationships within the data by applying techniques such as Association Rule Mining, leading to improved disease prediction, patient care, and resource allocation (Domadiya & Rao, 2019). These techniques are instrumental in identifying trends and associations that may need to be evident through conventional analysis methods.

Focusing on a specific instance, this study utilizes anonymized patient visit data from Hospital X, which includes over 65,535 patient entries from 2019 to 2023, coded with ICD-10. This dataset provides a rich source of information for exploring the complex interrelations of diseases within the hospital's patient population. By applying data mining techniques, particularly Association Rule Mining, this research aims to decode the intricate web of disease patterns, offering insights that could influence future healthcare decisions and strategies in Indonesia (Chaudhry et al., 2023).

In advance, by integrating advanced data mining techniques with robust classification systems like ICD-10, the research at Hospital X represents a significant step forward in understanding and managing health in Indonesia. Through this study, we aim to highlight the potential of data-driven approaches in enhancing the efficacy and efficiency of healthcare services nationwide (Cascini et al., 2021).

The research aims to delve into Hospital X's data using Association Rule Mining, seeking hidden disease patterns that regular epidemiological studies might miss. This exploration is crucial for predictive health analytics and strategic healthcare planning. Furthermore, the study intends to assess Association Rule Mining's effectiveness in leveraging ICD-10 coded data to gain deeper insights into disease trends and improve patient management strategies within Indonesia's healthcare landscape. Ultimately, this research strives to connect extensive data resources with practical healthcare insights, employing advanced data mining methods to enhance Indonesia's healthcare system's responsiveness and knowledge base (Karatat et al., 2022).

METHOD

This section outlines the methodology for researching "Application of Data Mining Techniques in Healthcare: Identifying Inter-Disease Relationships through Association Rule Mining" using the data from Hospital X. The methodology is structured as follows:

Data Collection

The data for this study consists of anonymized patient visit records from Hospital X, spanning from 2019 to 2023. Each record includes the visit date and the diagnosis codes assigned according to the ICD-10 classification. The dataset encompasses 65,535 entries, reflecting diverse patient interactions within the hospital.

Data Preprocessing

Before applying any data mining techniques, the dataset will undergo several preprocessing steps:

1. **Data Cleaning:** Remove any inconsistencies or errors in the data, such as duplicate records or missing values.

2. Data Transformation: Convert the diagnosis codes into a binary matrix format where rows represent patient visits and columns represent ICD-10 codes. Each entry in the matrix will be set to 1 if the diagnosis was made during the visit and 0 otherwise.
3. Data Reduction: Filter out infrequent ICD-10 codes to focus on the most common and potentially more relevant ones for discovering significant associations.

Application of Association Rule Mining

Association Rule Mining will be implemented to discover exciting relationships between diseases:

1. Algorithm Selection: Use the FP-Growth algorithm, which is efficient for datasets where specific patterns frequently occur.
2. Setting Parameters: Define the minimum support and confidence thresholds to identify meaningful rules. The support threshold will help eliminate rare itemsets. In contrast, the confidence threshold will ensure that only associations strong enough to be considered reliable are included.
3. Rule Generation: Generate association rules from the frequent itemsets discovered by the FP-Growth algorithm. These rules will indicate which diseases tend to co-occur within the dataset.

Analysis of Results

The generated rules will be analyzed to identify and interpret significant disease relationships:

1. Rule Evaluation: Assess the rules based on their support, confidence, and lift values. Rules with high lift values are exciting as they suggest a strong positive association between the antecedent and the consequent.
2. Medical Relevance: Consult with medical experts to interpret the findings in the context of clinical significance and existing medical knowledge.

Here is the flowchart illustrating the research methodology that will be implemented:

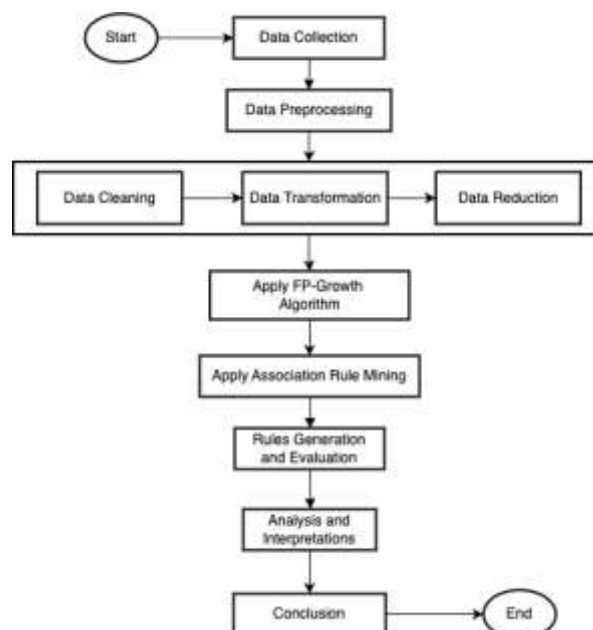


Figure 1. Research Flowchart

RESULTS AND DISCUSSION

Data Collection

The initial phase of our research involved a comprehensive examination of the raw data collected from Hospital X. This dataset encompasses patient visit records from 2019 to 2023, with each entry detailing the visit date and the respective ICD-10 diagnosis codes. The raw data provides a foundational understanding of the patient demographics, frequency of visits, and the diversity of diagnoses made during the period under review (Ng et al., 2016).

Below is a table representing a snapshot of the raw data. This table includes sample entries from the dataset, illustrating how the data is structured and the type of information available for each patient visit.

Table 1. Raw Data

Visit Number	Medrec	Date of visit	Diagnosis Code
RJ2019F13503	051979	01-Dec-19	R50
RI2019F02696	029924	13-Oct-19	D64
RJ2021G75803	037799	13-Dec-21	Z50.1
.....			
RJ2019F20067	004145	31-Dec-19	R50

Source : Analysis Results 2024

This initial analysis is crucial for identifying any discrepancies, missing values, or outliers in the data that may affect subsequent data processing and analysis stages. By ensuring the integrity and completeness of the raw data, we establish a robust basis for applying data mining techniques to uncover meaningful patterns and associations between diseases.

Data Preprocessing

Following the initial review of the raw data, a pivotal transformation was conducted to analyze the patterns in the patient visits further. This transformation involved creating a pivot table where the rows represent the dates of patient visits and the columns correspond to the ten most frequently occurring diseases based on the ICD-10 diagnosis codes. This restructuring allows for an aggregated data view, highlighting primary health conditions' prevalence and temporal distribution over the specified period.

The pivot table created for this study includes the ten most frequently occurring diseases among the patients visiting Hospital X from 2019 to 2023. These diseases, represented by their respective ICD-10 codes, are pivotal in understanding the health trends and challenges within the patient community served by the hospital. Below is a brief description of each disease based on the ICD-10 classification:

Table 2. Frequencies of 10 Most Common Diseases

No.	ICD-10 Code	ICD Description	Frequencies
1	Z09.8	Follow-up exam after other treatment for other conditions	9383
2	I10	Essential (primary) hypertension	4102
3	E11	Non-insulin-dependent diabetes mellitus	3761
4	I63.9	Cerebral infarction, unspecified	2675
5	R50.9	Fever, unspecified	2474
6	K30	Dyspepsia	1687
7	H26	Another cataract	1384
8	A91	Dengue hemorrhagic fever	1318
9	A01.0	Typhoid fever	1284

No.	ICD-10 Code	ICD Description	Frequencies
10	M54.5	Low back pain	1191

Source : Analysis Results 2024

Below is the pivot table that displays the date of visit along with the counts of the ten most common diseases:

Table 3. Transformed Data with Ten Most Common Diseases

Date	Z09.8	I10	E11	I63.9	R50.9	K30	H26	A91	A01.0	M54.5
2019-09-30	0	0	0	0	5	0	0	3	6	0
2019-10-01	0	0	0	0	1	0	0	0	2	0
2019-10-02	0	0	0	0	1	1	0	1	1	0
2019-10-03	0	0	0	0	3	2	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...
2023-07-06	0	1	0	0	0	0	0	0	0	0

Source : Analysis Results 2024

Following the creation of the pivot table, the next crucial step involved transforming the data to prepare for Association Rule Mining. This transformation entailed converting the dataset into a binary format, where each cell in the table was assigned a value based on the presence or absence of a disease diagnosis on a particular visit date. Specifically, entries with a recorded diagnosis were marked with '1', indicating the presence of the disease. In contrast, cells without a diagnosis were marked with '0', indicating its absence.

This binary transformation is essential for the application of Association Rule Mining, as it simplifies the data and focuses the analysis on the occurrence and association of diseases rather than their frequency. This format allows the mining algorithm to efficiently identify and generate rules based on the co-occurrence of diseases across different patient visits.

Below is the transformed table, now formatted to suit the requirements of the Association Rule Mining process:

Table 4. Data Transformation for Data Mining Implementation

Date	Z09.8	I10	E11	I63.9	R50.9	K30	H26	A91	A01.0	M54.5
2019-09-30	0	0	0	0	1	0	0	1	1	0
2019-10-01	0	0	0	0	1	0	0	0	1	0
2019-10-02	0	0	0	0	1	1	0	1	1	0
2019-10-03	0	0	0	0	1	1	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...
2023-07-06	0	1	0	0	0	0	0	0	0	0

Source : Analysis Results 2024

This binary format highlights patterns and relationships that might need to be more evident when considering the raw count data. It sets the stage for the subsequent analysis phase, which involves applying the Association Rule Mining algorithm to discover significant associations between the diseases that frequently co-occur during patient visits.

Results of Data Mining Implementation Using FP-Growth and Association Rules Mining

The application of the FP-Growth algorithm followed by Association Rule Mining has yielded significant insights into the relationships between various diseases recorded in the data from Hospital X. By setting the minimum support at 0.95 and the minimum confidence at 0.9; we ensured that only the most relevant and frequently occurring disease associations were considered, thereby

focusing on the most impactful relationships. The results table, which will be provided below, details the calculated data from this analysis:

Table 5. FP-Growth and Association Rule Results

No	Premises	Conclusion	Support	Confidence	Laplace	Gain	p-s	Lift	Conviction
10	K30, A01.0, A91	R50.9	0.388	0.911	0.973	-0.464	0.061	1.185	2.604
11	A91, I63.9	R50.9	0.314	0.911	0.977	-0.375	0.049	1.185	2.608
12	I10, A91	R50.9	0.413	0.913	0.973	-0.491	0.065	1.188	2.662
13	M54.5	K30	0.432	0.914	0.972	-0.513	0.066	1.180	2.615
14	M54.5	I10	0.432	0.914	0.972	-0.513	0.099	1.296	3.425
15	K30, I10, M54.5	I63.9	0.368	0.917	0.976	-0.435	0.170	1.860	6.087
16	I10, A01.0, A91	R50.9	0.362	0.919	0.977	-0.426	0.059	1.195	2.847
17	K30, R50.9, I10, M54.5	I63.9	0.314	0.919	0.979	-0.369	0.146	1.865	6.274
18	K30, I10, A91	R50.9	0.368	0.920	0.977	-0.432	0.060	1.197	2.889
19	A91, E11	R50.9	0.318	0.920	0.979	-0.374	0.052	1.197	2.895
20	I10, A91, I63.9	R50.9	0.304	0.921	0.980	-0.356	0.050	1.197	2.914
21	A01.0, I63.9	K30	0.355	0.921	0.978	-0.416	0.056	1.189	2.843
22	I10, A01.0, I63.9	K30	0.340	0.921	0.979	-0.398	0.054	1.189	2.861
23	I10, I63.9	K30	0.426	0.921	0.975	-0.499	0.068	1.189	2.866
24	K30, I10, A01.0, A91	R50.9	0.328	0.922	0.980	-0.384	0.055	1.200	2.980
25	I10, A01.0, M54.5	I63.9	0.317	0.924	0.981	-0.369	0.148	1.875	6.651
26	R50.9, I63.9	K30	0.376	0.925	0.978	-0.438	0.061	1.194	3.004
27	R50.9, I10, I63.9	K30	0.363	0.926	0.979	-0.422	0.059	1.195	3.041
28	R50.9, I63.9, M54.5	K30, I10	0.314	0.927	0.982	-0.363	0.114	1.567	5.598
29	R50.9, I10, A01.0, I63.9	K30	0.299	0.928	0.982	-0.346	0.049	1.198	3.126
30	I10, A01.0, M54.5	K30	0.318	0.928	0.982	-0.368	0.053	1.198	3.128
31	A91, I63.9	K30	0.320	0.928	0.982	-0.369	0.053	1.198	3.141
32	A01.0, M54.5	K30	0.342	0.929	0.981	-0.394	0.057	1.199	3.167
33	K30, M54.5	I10	0.401	0.929	0.979	-0.462	0.097	1.318	4.173
34	I10, M54.5	K30	0.401	0.929	0.979	-0.462	0.067	1.200	3.186
35	R50.9, A01.0, I63.9	K30	0.310	0.930	0.983	-0.356	0.052	1.201	3.224
36	K30, A01.0, M54.5	I10	0.318	0.932	0.983	-0.365	0.078	1.322	4.334
37	A01.0, M54.5	I10	0.343	0.933	0.982	-0.392	0.084	1.323	4.391
38	R50.9, M54.5	K30	0.363	0.933	0.981	-0.416	0.062	1.204	3.354
39	I63.9, E11	K30	0.304	0.933	0.984	-0.347	0.052	1.204	3.364
40	I10, A91, I63.9	K30	0.308	0.934	0.984	-0.352	0.053	1.206	3.409
41	R50.9, I10, M54.5	K30	0.342	0.936	0.983	-0.388	0.059	1.209	3.534
42	R50.9, M54.5	I10	0.365	0.937	0.982	-0.414	0.090	1.329	4.652
43	I63.9	I10	0.462	0.938	0.980	-0.523	0.115	1.331	4.763
44	I63.9, M54.5	K30	0.381	0.939	0.982	-0.430	0.067	1.212	3.697
45	K30, R50.9, M54.5	I10	0.342	0.940	0.984	-0.385	0.085	1.333	4.918
46	A01.0, I63.9, M54.5	K30	0.307	0.942	0.986	-0.344	0.054	1.216	3.882
47	I10, I63.9, M54.5	K30	0.368	0.948	0.985	-0.408	0.067	1.223	4.297

No	Premises	Conclusion	Support	Confidence	Laplace	Gain	p-s	Lift	Conviction
48	R50.9, I63.9, M54.5	K30	0.321	0.948	0.987	-0.356	0.059	1.224	4.374
49	R50.9, I10, I63.9, M54.5	K30	0.314	0.952	0.988	-0.346	0.058	1.228	4.649
50	K30, I63.9	I10	0.426	0.954	0.986	-0.467	0.111	1.354	6.470
51	I63.9, M54.5	I10	0.388	0.957	0.988	-0.423	0.102	1.358	6.860
52	A91, I63.9	I10	0.330	0.958	0.989	-0.359	0.087	1.359	6.993
53	A01.0, I63.9	I10	0.369	0.958	0.988	-0.401	0.098	1.360	7.108
54	K30, A01.0, I63.9	I10	0.340	0.959	0.989	-0.369	0.090	1.360	7.199
55	K30, A91, I63.9	I10	0.308	0.964	0.991	-0.331	0.083	1.367	8.114
56	R50.9, I63.9	I10	0.392	0.964	0.990	-0.422	0.106	1.368	8.262
57	I63.9, E11	I10	0.314	0.964	0.991	-0.337	0.084	1.368	8.262
58	K30, R50.9, I63.9	I10	0.363	0.965	0.990	-0.390	0.098	1.369	8.491
59	K30, I63.9, M54.5	I10	0.368	0.966	0.991	-0.394	0.099	1.370	8.589
60	K30, R50.9, A01.0, I63.9	I10	0.299	0.967	0.992	-0.320	0.081	1.372	8.978
61	R50.9, A91, I63.9	I10	0.304	0.968	0.992	-0.324	0.082	1.373	9.105
62	R50.9, A01.0, I63.9	I10	0.323	0.969	0.992	-0.343	0.088	1.375	9.653
63	A01.0, I63.9, M54.5	I10	0.317	0.973	0.993	-0.334	0.087	1.381	11.016
64	R50.9, I63.9, M54.5	I10	0.330	0.974	0.993	-0.347	0.091	1.382	11.458
65	K30, R50.9, I63.9, M54.5	I10	0.314	0.977	0.994	-0.328	0.088	1.386	13.042

Source : Analysis Results 2024

The results are presented in a detailed table that lists the premises, conclusions, and metrics such as support, confidence, Laplace, gain, p-s, lift, and conviction values for each rule identified. This tabular data quantitatively measures the strength and reliability of each association rule discovered during the analysis.

For example, one of the most vital rules discovered (Rule No. 65) shows that the presence of conditions coded as K30 (Dyspepsia), R50.9 (Fever, unspecified), I63.9 (Cerebral infarction, unspecified), and M54.5 (Low back pain) together in a patient significantly increases the likelihood of a diagnosis of I10 (Essential hypertension), with a confidence of 97.7% and a lift of over 13, indicating an influential association that is likely not due to random chance.

This analysis highlights the common co-occurrences and points towards potential comorbid conditions that could inform clinical decisions and healthcare policy planning. Below is a graphical representation of the association rules derived from the analysis, showing the complex interplay between various diseases:



Figure 1. Isometric Graph of Diseases and Rules

This visualization helps in quickly understanding the strength and direction of the relationships among different ICD-10 coded diseases, providing a clear picture of how multiple conditions correlate within the patient population of Hospital X. The graphical representation serves as a powerful tool for communicating the results to stakeholders involved in healthcare management and planning.

This study's use of Association Rule Mining (ARM) with the FP-Growth algorithm to identify inter-disease relationships aligns with existing research such as Sanati-Mehrziy et al. and Kulkarni and Mundhe, demonstrating ARM's effectiveness in healthcare data analysis. Dicuonzo et al. emphasize the growing importance of big data in healthcare, a trend reflected in this study's approach to digital healthcare data, supported by the focus on electronic medical records (EMR) highlighted by (Berros et al., 2023)

Significant findings from this study are consistent with those of Wang et al., who noted the effectiveness of sophisticated data mining techniques in healthcare (Sun et al., 2018). The accuracy of medical coding discussed by the Thought Leadership Team ensures the reliability of ARM in analyzing disease codes, enhancing the applicability of the findings to public health surveillance and interventions, as noted by (Morgenstern et al., 2021)

Furthermore, the research supports the call for diversity and inclusive leadership in healthcare (Ashikali et al., 2021), as understanding disease interactions can lead to better health outcomes by tailoring healthcare practices to diverse patient needs. Overall, this study reinforces the value and feasibility of using data mining to enhance healthcare management and decision-making, contributing valuable insights to the field and promoting a more informed approach to healthcare policy.

CONCLUSION

This study utilized the FP-Growth algorithm and Association Rule Mining to analyze patient data from Hospital X, uncovering significant inter-disease relationships. The findings highlight strong associations between diseases like hypertension and diabetes, confirming the utility of data mining in healthcare for enhancing predictive modeling and resource allocation. These insights can inform future healthcare strategies, emphasizing integrating data analytics into clinical practice for

improved patient outcomes and more personalized treatment plans. This approach encourages a data-driven direction in healthcare, aligning with global trends towards innovative solutions.

REFERENCES

- Alonge, O., Sonkarlay, S., Gwaikolo, W., Fahim, C., Cooper, J. L., & Peters, D. H. (2019). Understanding the role of community resilience in addressing the Ebola virus disease epidemic in Liberia: a qualitative study (community resilience in Liberia). *Global Health Action*, 12(1), 1662682.
- Alyahya, M. S., & Khader, Y. S. (2019). Health care professionals' knowledge and awareness of the ICD-10 coding system for assigning the cause of perinatal deaths in Jordanian hospitals. *Journal of Multidisciplinary Healthcare*, 149–157.
- Ashikali, T., Groeneveld, S., & Kuipers, B. (2021). The role of inclusive leadership in supporting an inclusive climate in diverse public sector teams. *Review of Public Personnel Administration*, 41(3), 497–519.
- Behzadnia, B., Deci, E. L., & DeHaan, C. R. (2020). Predicting relations among life goals, physical activity, health, and well-being in elderly adults: a self-determination theory perspective on healthy aging. *Self-Determination Theory and Healthy Aging: Comparative Contexts on Physical and Mental Well-Being*, 47–71.
- Berros, N., El Mendili, F., Filaly, Y., & El Bouzekri El Idrissi, Y. (2023). Enhancing digital health services with big data analytics. *Big Data and Cognitive Computing*, 7(2), 64.
- Cascini, F., Santaroni, F., Lanzetti, R., Failla, G., Gentili, A., & Ricciardi, W. (2021). Developing a data-driven approach in order to improve the safety and quality of patient care. *Frontiers in Public Health*, 9, 667819.
- Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023). A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective. *Symmetry*, 15(9), 1679.
- Domadiya, N., & Rao, U. P. (2019). Privacy Preserving Distributed Association Rule Mining Approach on Vertically Partitioned Healthcare Data. *Procedia Computer Science*, 148, 303–312. <https://doi.org/https://doi.org/10.1016/j.procs.2019.01.023>
- Groseclose, S. L., & Buckeridge, D. L. (2017). Public health surveillance systems: recent advances in their use and evaluation. *Annual Review of Public Health*, 38, 57–79.
- Gul, S., Bano, S., & Shah, T. (2021). Exploring data mining: facets and emerging trends. *Digital Library Perspectives*, 37(4), 429–448.
- Hilbert, M. (2016). Big data for development: A review of promises and challenges. *Development Policy Review*, 34(1), 135–174.
- Karatas, M., Eriskin, L., Deveci, M., Pamucar, D., & Garg, H. (2022). Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives. *Expert Systems with Applications*, 200, 116912.
- Morgenstern, J. D., Rosella, L. C., Daley, M. J., Goel, V., Schünemann, H. J., & Piggott, T. (2021). “AI’s gonna have an impact on everything in society, so it has to have an impact on public health”: a fundamental qualitative descriptive study of the implications of artificial intelligence for public health. *BMC Public Health*, 21, 1–14.
- Ng, K., Steinhubl, S. R., DeFilippi, C., Dey, S., & Stewart, W. F. (2016). Early detection of heart failure using electronic health records: practical implications for time before diagnosis, data diversity, data quantity, and data density. *Circulation: Cardiovascular Quality and Outcomes*, 9(6), 649–658.
- Senbekov, M., Saliev, T., Bukeyeva, Z., Almaybayeva, A., Zhanaliyeva, M., Aitenova, N., Toishibekov, Y., & Fakhradiyev, I. (2020). The recent progress and applications of digital technologies in healthcare: a review. *International Journal of Telemedicine and Applications*, 2020.

- Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., & Wang, G. (2018). Data processing and text mining technologies on electronic medical records: a review. *Journal of Healthcare Engineering*, 2018.
- Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13.



© 2024 by the authors. It was submitted for possible open-access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).