Section: Miscellaneous

Original Research Article

ASSOCIATION RULE MINING AND STATISTICAL VALIDATION FOR PROGNOSIS OF INFECTIOUS DISEASES IN KATHMANDU, NEPAL

Sarad Chandra Kafle[1], Mohit Nain[2], Ramendra Kumar Raman[3], Hom Prasad Adhikari[4]

[1]*Research Scholar, Department of Statistics, Om Sterling Global University, Hisar, India*
[2]*Research Supervisor, Department of Statistics, Om Sterling Global University, Hisar, India*
[3]*Associate Professor, Clinical Research, Dr KNS Memorial Institute of Medical Sciences, Barabanki, India*
[4]*Director, Suvekchya International Hospital, Kathmandu, Nepal*

Received : *03/09/2025*
Received in revised form : *17/10/2025*
Accepted : *05/11/2025*

Corresponding Author:
Sarad Chandra Kafle,
Research Scholar, Department of Statistics, Om Sterling Global University, Hisar, India.
Email: sarad.kafle49@gmail.com

DOI: 10.70034/ijmedph.2025.4.234

Source of Support: Nil,
Conflict of Interest: None declared

Int J Med Pub Health
2025; 15 (4); 1306-1313

ABSTRACT

Background: Cities like Kathmandu, Nepal, witness rapid urbanization and population growth, which increases the transmission of infectious diseases and presents considerable public health challenges. .

Materials and Methods: This study employs retrospective method of data collection from Sukraraj Tropical and Infectious Diseases Hospital in Teku, Kathmandu with nine hundred and sixty nine (n=968) patient's records, and used data mining technique of Association Rule Mining to uncover hidden patterns and relationships in inpatients' records focusing on dengue, scrub typhus and kala-azar. The association rules provided by Apriori algorithm are validated using Chi-square test and Bootstrapping.

Results: Association mining algorithm provided nine, seven and eight interesting association rule for scrub typhus, dengue and kala-azar respectively. These association rule demonstrated that patients with moderate length of stay and no comorbid conditions, specifically males with dengue and females with scrub typhus, had high discharge rates.

Conclusion: Findings of this study demonstrates the important connection between outcomes of patients and their demographics, however only the association rules having higher lift are statistically validated. The approach and findings can be used to guide strategies focused on public health, disease control, and the allocation of health care resources.

Keywords: Association rule, data mining, infectious disease, prognosis.

INTRODUCTION

Association Rule Mining (ARM) is a vital data mining technique for revealing unseen patterns and correlations in large datasets. It focuses on the discovery of associations and the frequent itemsets which are crucial for insight-based decision making and strategic Planning. In the field of Health Informatics ARM has its importance in pattern extraction from medical data during knowledge discovery, enhancing the ability to predict diseases.[1] ARM is vital for predicting health-related events, revealing implicit relationships between variables and for the analysis of complex multidimensional problems, such as infectious diseases.

Infectious disease represents a major worldwide public health challenge, particularly for developing countries such as Nepal. The capital, Kathmandu, has a growing burden of infectious diseases due to rapid urbanization, environmental changes, and high population density.[2] One of the principal public health challenges in the city is the public health threat posed by infectious diseases, including respiratory and gastrointestinal diseases and vector-borne diseases such as dengue and scrub typhus.[3] For strategic public health resource allocation and planning, predicting the occurrence and consequences of the diseases is pivotal. In addition, the detailed examination of the relationships between variables using association rule mining is a crucial technique for guiding decision-makers in the planning and implementation stages. However, the use of data mining to analyze disease outcomes and relationships for infectious diseases in Kathmandu is sorely lacking. This study aims to fill this gap by using Association Rule Mining (ARM) to identify the associative relationships between epidemiological variables and infectious diseases in Kathmandu, Nepal.

This study primarily seeks to assess the use of ARM techniques to identify patterns in infectious disease registry data and analyze the associations between demographic and other factors concerning the outcomes of patients hospitalized with dengue and scrub typhus.

Potential uses of data mining for disease prognosis have been documented in the literature. Abeysinghe and Cui (2018),[4] worked on Query-Constraint-Based Mining (QARM) on clinical data and dealt with some of the problems posed by categorical data. They demonstrated that the incorporation of QARM into web systems improves exploratory data analysis by filtering out useless patterns and emphasizing valuable pieces of information. In the same vein, Ghorbani and Ghousia (2019),[5] focused on predictive data mining techniques in medical diagnosis and underlined the importance of their utilization in the Kathmandu, Nepal region where powerful predictive tools for infectious diseases are highly needed. This study documented the use of data mining in the healthcare sector and the gaps that remain and advocated for the use of advanced analytical methods as the healthcare sector continues to generate large volumes of data. Ostapchuk (2019),[6] introduced an interactive pattern mining framework for electronic health records (EHR) that exposes disease-comorbidity and disease-gene correlations. This study brought forth the 4P medicine model—predictive, preventive, personalized, and participatory—and the author hoped that the combination of innovative data mining techniques and classical approaches to statistics would encourage improvement in patient care, although he acknowledged the need for more innovative methods that would focus on real-world problems to be solved.

Association rule mining's Apriori algorithm has been used in a variety of fields including the forecasting and diagnosing of heart disease,[7,8] brain tumor,[9] dengue fever,[10] chronic,[11] and infectious diseases,[12] and drug resistant in tuberculosis.[13] Its use is particularly important with regards to COVID-19,[14,15] paediatric primary care decision support systems,[16] and emergency department patients' laboratory diagnostic tests.[17] This paper employed ARM to discover interesting association rules involving patients' demographic data, duration of hospital stay, and outcomes of a hospital stay. This analysis aims to use ARM to enhance understanding of the relationships concerning the different factors implicated with dengue and scrub typhus. The author not being a health professional is the reason factors such as medical history and treatment which could have been important were not studied.

MATERIALS AND METHODS

This retrospective study utilized the inpatient records at Sukraraj Tropical and Infectious Disease Hospital, Tehku, Kathmandu. Ethical approval was obtained from the Nepal Health Research Council, and data collection with the consent of the hospital. The dataset included the records of 969 admissions from the hospital, with patients being admitted from 16th of February 2021 to the 21st of March 2024, and having one of the following diagnoses: dengue, scrub typhus, or kala-azar. The hospital is the only specialized clinic for infectious diseases in Kathmandu and receives a majority of such cases. This is why we selected this hospital.

The data was analyzed through Association Rule Mining (ARM) applying the Apriori algorithm in R Studio (Windows). The objective was to determine patterns of the factors affecting the outcomes of patients. Each rule was assessed using support, confidence, and lift which measured the frequency, reliability, and association strength, respectively.

The dataset encompassed both demographic and clinical variables: sex (male, female), age, duration of hospital stay, presence of plural disease (yes or no), and outcome. The age categories were child (younger than 18), young adult (18–30), middle adult (31–50), older adult (51–65), and elderly (over 65).[18] The duration of hospital stay was classified as short (4 days or fewer), moderate (5–9 days), and prolonged (10 days or more).[19] The outcomes collected from hospital records—discharged, referred, leave against medical advice (LAMA), discharge on patient's request (DOPR), and expired—were recoded to reflect prognosis as favorable (discharged), indeterminate (LAMA, DOPR, referred), and unfavorable (expired).[20]

Apriori Algorithm

The approach proposed by Agrawal et al. (1993),[21] begins with generating frequent itemsets one item at a time. It then recursively generates frequent sets with two items, then three, and so on, until generating all sizes of frequent itemsets.[22] The strength of the rules generated from the Apriori Algorithm are assessed based on support, confidence, and lift ratio.[22] While discussing the strength of an association rule, the terms antecedent and consequent came frequently. Antecedent denotes the IF part and consequent denotes the THEN part in the co-occurrence of itemsets. Confidence is the ratio of the transactions that contain all antecedent and consequent itemsets to the transactions that contain all the antecedent itemsets. They are referred to as the support of the association rule. This can be displayed mathematically as defined by Shmueli et al. (2018),[22] as follows:

$$\text{Support} = \frac{\text{Number of cases antecedent and consequent occur together}}{\text{total number of transactions}}$$

= P (antecedent AND consequent) .. (i)

Also,

Confidence =

$$\frac{\text{Number of transactions with both antecedent and consequent itemsets}}{\text{Number of transactions with antecedent itemsets}} \quad \ldots\ldots \text{(ii)}$$

$$= \frac{\text{P(antecedent AND consequent)}}{\text{P(antecedent)}}$$

= P (consequent/antecedent)

Thus, confidence is the conditional probability of a consequent given a case of an antecedent. Another measure for the strength of an association rule is the lift ratio, which is the expected confidence value if the antecedent and consequent itemsets are independent.

$$\text{Lift ratio} = \frac{Confidence}{Benchmark\ confidence} \quad \ldots\ldots \text{(iii)}$$

Where,

Benchmark confidence =

$$\frac{\text{number of transactions with consequent itemset}}{\text{number of transactions in database}} \ldots \text{(iv)}$$

In this study, strength of association rule were evaluated by comparing the value of support, confidence, and lift. Confidence, which indicates how frequently a rule holds, is the conditional probability of the consequent given the antecedent. Nevertheless, even in the case of independent occurrences, confidence may suggest a rule is strong if both the antecedent and consequent are frequent on their own. In response, we calculate lift; a measure of how much more frequent the antecedent and consequent are together in relation to how frequently independent (not correlated) pairs would be. Mathematically, lift ($A \Rightarrow B$) = Confidence ($A \Rightarrow B$) / Support (B) = $P(A \wedge B) / [P(A) \times P(B)]$ and a lift greater than 1 indicates a meaningful positive association. The higher the lift, the more substantial and more statistically significant the relationship between antecedent and consequent.[22]

Statistical validation of association rules

To ensure that the association rules derived through the Apriori algorithm represented meaningful patterns rather than spurious correlations, two complementary validation procedures were applied: (i) the Chi-square test of independence, and (ii) a non-parametric bootstrap procedure.

1. Chi-square Test of Independence

For each association rule generated, a contingency table was constructed between the antecedent (LHS) and consequent (RHS) variables using the original dataset. The Chi-square test of independence was then applied to examine whether there was a statistically significant association between the two sets of variables.

The test statistic is defined as:

$$\chi^2 = \sum \frac{(O_{ij}-E_{ij})^2}{E_{ij}} \text{---------- (V)}$$

Where Oij indicates associated frequency while Eij indicates expected frequency based on the null hypothesis on independence. Association rules with p-values from chi-square distribution lower than 0.05 are considered statistically significant. The literature recommends the use of Chi-square as it tests the very hypothesis of independence that must be widely accepted in order for association rules to be valid.[23,24] The Chi-square test is the only one that allows the theorist to make a statistical contrast as opposed to the support and confidence calculation which are purely descriptive.

2. Bootstrapping for Stability Assessment

A non-parametric bootstrap method[25] was conducted with 1,000 resamples. For each resample, recurrence frequencies of rules were noted after each execution of the Apriori algorithm. Any rules that reappeared in ≥ 70 % of resamples with stable support and confidence were considered stable.[26]

The two-stage validation mentioned above guaranteed that the final association rules were both non-parametric statistically significant and resampling robust. This is vital considering the imbalanced public-health data, where there are fewer indeterminate or adverse outcomes.

RESULTS

The sample population of inpatients selected for this study contains a diverse demographic composition with different hospital stay durations and outcomes after hospital stay. The Gender distribution shows that male patients (58.1%) are somewhat higher in proportion than female patients (41.9%) admitted to STIDH (see Figure 1). Age distribution of sample inpatients shows that their mean age is 37.7 years with a standard deviation of 16.3 years. The inpatient population ranged from 2 to 92 years of age (see Figure 2). Whereas most of the patients (87.8%) were discharged after certain days of hospital stay, having recovered from the disease, but some patients had LAMA (5.9%), and some were referred (4.5%) to another hospital for better treatment, only a few (0.8%) patients died (see Figure 3). These data show a good level of treatment at the study hospital. Duration distribution of inpatients shows a positively skewed distribution with a mean of 4.2 days and a standard deviation of 3.7 days, where the patients were admitted for a minimum of one day to a maximum of 36 days to recover from the disease (see Figure 4). The distribution of diseases among patients indicates a notable dominance of Dengue cases, with 771 (85.4%) patients. This suggests that Dengue is the most common infectious disease among hospitalized patients during the study period. Following Dengue, the next most frequently observed diseases are Kala-azar and Scrub Typhus, accounting for 3.6% and 3.2% of admissions, respectively. A small number of patients experienced multiple infections, including Dengue with Scrub Typhus (2.7%) and Covid-19 with Dengue (1.0%). Additionally, combinations such as Dengue with Kala-azar and COVID-19 with Scrub Typhus each represent 0.2% of total admissions (see Figure 5). The Treemap visualization illustrates that a significant majority of patients were from Kathmandu (52.7%), the district where the hospital is located. The next most common districts of origin were Lalitpur (8%), Dhading (7.9%), Bhaktapur (5%), and Nuwakot

(3.1%). These five districts collectively account for over 75% of all admitted patients, reflecting the hospital's central role in providing healthcare services to the Kathmandu Valley and its surrounding districts [Figure 6].

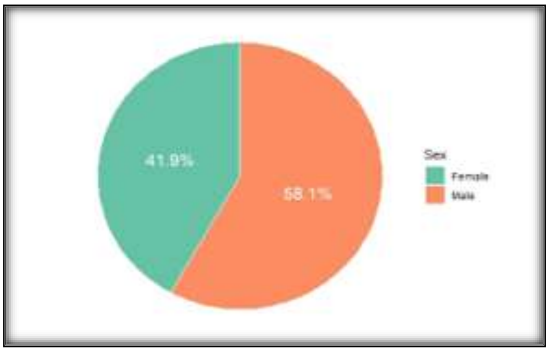

Figure 1: Sex-wise Distribution of Patients

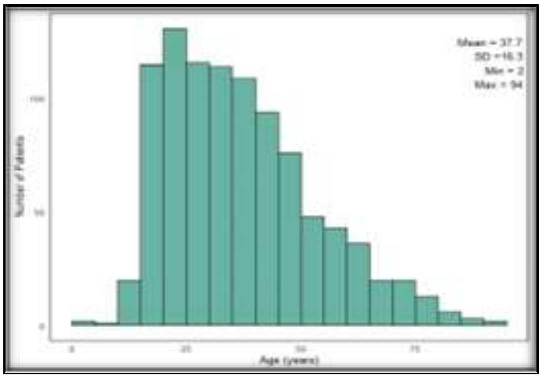

Figure 2: Age Distribution of Patients

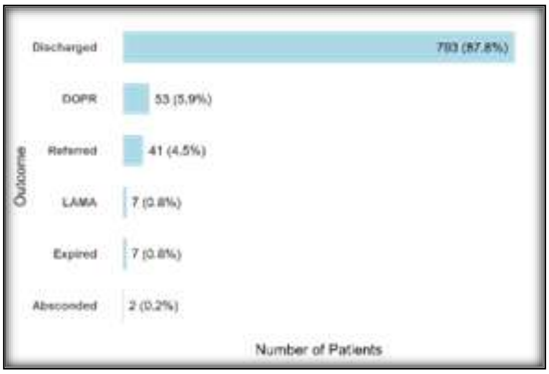

Figure 3: Treatment Outcomes of Admitted Patients

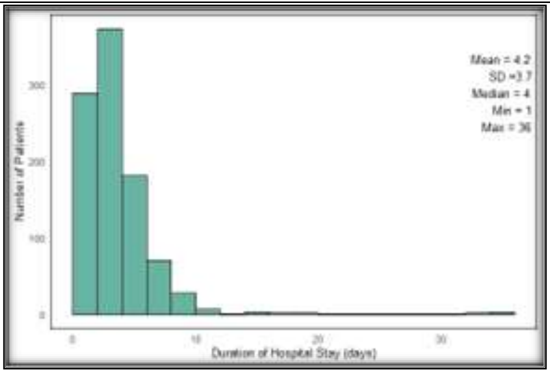

Figure 4: Hospital Stay Duration Distribution

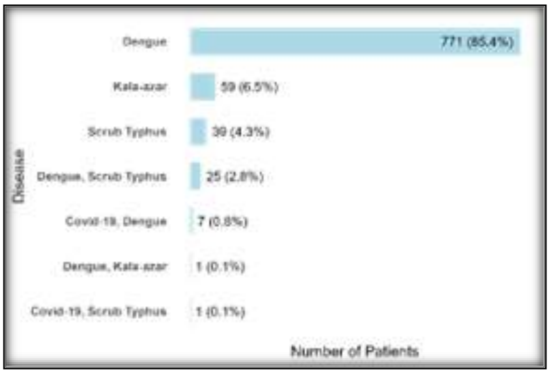

Figure 5: Distribution of Diagnosed Infectious Diseases

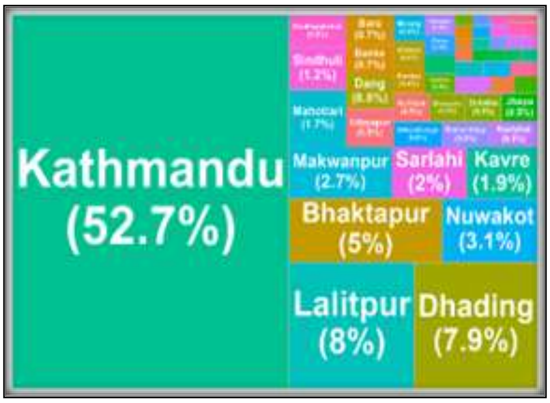

Figure 6: Geographical Distribution of Patients by District

ARM applied for Dengue (n=804), scrub typhus (n=65) and kala-azar (n=60) inpatient data and only the interesting association rules are tabulated below:

Table 1: Association Rules for Scrub typhus Cases, test of association using Fisher exact and Bootstrapping (n=65)

SN	Association rules	Support	Confidence	Lift	P-value	Bootstrap		
						Support	Confidence	Percentage
1	{District=Dhading}→{Prognosis=Favorable}	0.2462	1	1.0833	0.3219	0.2443 ±0.0551	1±0.00	100
2	{Age group=45–59 years}→{Prognosis=Favorable}	0.2	1	1.0833	0.5739	0.2011±0.0516	1±0.00	99.8
3	{Caste=Brahmin, Plural=Yes}→{Prognosis=Favorable}	0.2154	1	1.0833	0.9999	0.2136±0.0503	1±0.00	100
4	{Sex=Female, Caste=Brahmin}→{Prognosis=Favorable}	0.1846	1	1.0833	0.9999	0.1836±0.0488	1±0.00	99.8
5	{Age group=18–35 years, Duration=Less than 5 days, Plural=No}→{Prognosis=Favorable}	0.1076	1	1.0833	0.9999	0.1168±0.0355	1±0.00	93

6	{Sex=Female, Duration=Less than 5 days, Plural=No}→{Prognosis=Favorable}	0.1077	1	1.0833	0.9999	0.1134±0.0367	1±0.00	94.4
7	{Sex=Female, Duration=5–10 days, Caste=Brahmin}→{Prognosis=Favorable}	0.1077	1	1.0833	0.9999	0.1144±0.0349	1±0.00	91.4
8	{Duration=Less than 5 days, Caste=Brahmin, Plural=Yes}→{Prognosis=Favorable}	0.1231	1	1.0833	0.9999	0.1275±0.0369	1±0.00	95.4
9	{Duration=Less than 5 days, District=Dhading}→{Prognosis=Favorable}	0.1385	1	1.0833	0.9999	0.14±0.043	1±0.00	98

Findings show that nine interesting association rules for the prognosis of scrub typhus were extracted, and all are associated with a favorable prognosis. Table 1 shows that these rules consistently show that early discharge from the hospital, especially within the first 5 days, plays a key role in determining prognosis, along with socio-demographic factors like age, sex, caste, and region. The rule {District=Dhading} → {Prognosis=Favorable} (support = 0.246, lift=1.0833) shows that patients from Dhading have a 24.6% chance of a Favorable prognosis, with an 8.33% higher likelihood of recovery compared to other regions. Similarly, {Age Group=45–59 years} → {Prognosis=Favorable} (support = 0.2, lift=1.0833) suggests that individuals in this age group are 8.33% more likely to recover favorably compared to others. Other notable rules include {Caste=Brahmin, Plural=Yes} → {Prognosis=Favorable} (support = 0.215, lift=1.0833), where Brahmin patients with multiple symptoms consistently show a positive outcome, and {Sex=Female, Caste=Brahmin} → {Prognosis=Favorable} (support = 0.185, lift=1.0833), indicating gender and caste may influence prognosis, possibly due to socio-economic factors. Rules involving young adults, such as {Age Group=18–35 years, Duration=Less than 5 days, Plural=No} → {Prognosis=Favorable} (support = 0.107, lift=1.0833), emphasize the critical role of early detection in younger populations. Notably, all rules were validated by bootstrapping with high support and confidence (percentage of rules repetition is above 70%) (see Table 1).

Table 2: Association Rules for Dengue Cases, test of association using Chi-square and Bootstrapping (n=804)

SN	Association rules	Support	Confidence	Lift	P-value	Bootstrap Support	Bootstrap Confidence	Bootstrap Percentage
1	{Age group=45–59}→{Prognosis=Favorable}	0.1778	0.9167	1.0569	0.0461	0.1774±0.0135	0.9153±0.0219	100
2	{Age group=45–59, Plural=No}→{Prognosis=Favorable}	0.1045	0.9545	1.0995	0.0111	0.1038±0.0104	0.9533±0.023	100
3	{Sex=Female, Caste=Janajati, Plural=No}→{Prognosis=Favorable}	0.102	0.9213	1.0613	0.1158	0.1021(0.0116)	0.9211±0.0277	100
4	{Sex=Male, Duration=5–10 days}→{Prognosis=Favorable}	0.1443	0.9206	1.0604	0.0580	0.1434±0.0124	0.9206±0.0243	100
5	{Sex=Male, Caste=Brahmin}→{Prognosis=Favorable}	0.153	0.9461	1.0898	0.0041	0.1529±0.0131	0.9463±0.0201	100
6	{Duration=5–10 days}→{Prognosis=Favorable}	0.2363	0.909	1.0471	0.0420	0.2357±0.0154	0.9083±0.0199	100
7	{Duration=5–10 days, Plural=Yes}→{Prognosis=Favorable}	0.1119	0.9278	1.0687	0.0634	0.1117±0.0112	0.9258±0.0282	100

Seven rules associated with prognosis for dengue were identified, interestingly all are associated with a favorable outcome. Unlike Scrub typhus, some of these rules were found to be statistically significant (P-value<0.05). Findings show that the association rule {Age Group=45–59} → {Prognosis=Favorable} (support = 0.1778, lift=1.0995) shows that 17.78% of patients in this age group experience a Favorable prognosis, with an 9.95% higher chance of recovery compared to other age groups. The other rule {Sex=Female, Caste=Janajati, Plural=No} → {Prognosis=Favorable} (support = 0.102, lift=1.0613) suggest that gender, caste, and symptom severity play a role in recovery, with female Janajati (ethnic) patients with fewer symptoms showing a 6.13% higher chance of a favorable outcome. Additionally, {Sex=Male, Duration=5–10 days} → {Prognosis=Favorable} (support = 0.1443, lift=1.0604) indicates that males with this illness duration consistently recover favorably in moderate duration of hospital stay. These findings point to early intervention, gender, and socio-cultural factors as crucial factors in optimizing treatment strategies for Dengue patients.

Table 3: Association Rules for Kala-azar, test of association using Fisher exact and Bootstrapping (n=60)

SN	Association rules	Support	Confidence	Lift	P-value	Bootstrap Support	Bootstrap Confidence	Bootstrap Percentage
1	{Duration=Less than 5 days, Caste=Janajati}→{Prognosis=Favorable}	0.3	1	1.0526	0.5471	0.2965±0.061	1±0.00	100

2	{Age group=18–35 years, Duration=Less than 5 days}→{Prognosis=Favorable}	0.2667	1	1.0526	0.2667	0.2612±0.0553	1±0.00	100
3	{Age group=18–35 years, Duration=Less than 5 days, Plural=No}→{Prognosis=Favorable}	0.2667	1	1.0526	0.2667	0.2612±0.0553	1±0.00	100
4	{District=Kathmandu}→{Prognosis=Favorable}	0.2333	1	1.0526	0.2333	0.2312±0.0524	1±0.00	100
5	{Age group=18-35 years, Sex=Male, Duration=Less than 5 days}→{Prognosis=Favorable}	0.2333	1	1.0526	0.2333	0.2306±0.0534	1±0	100
6	{Sex=Male, Duration=Less than 5 days, Caste=Janajati}→{Prognosis=Favorable}	0.2333	1	1.0526	0.2333	0.4669±0.067	0.9326±0.0436	99.6
7	{Sex=Female}→{Prognosis=Favorable}	0.1833	1	1.0526	0.1833	0.1801±0.0497	1±0.00	100
8	{Duration=Less than 5 days, Caste=Janajati, Plural=No} → {Prognosis=Favorable}		1	1.0526	0.9999	0.2134±0.0548	1±0.00	99.8

Findings show that eight rules associate with favorable prognosis for Kala-azar patients (Table 3). All rules showed a Confidence of 1.00 and a Lift of 1.0526, suggesting factors leading to the outcome always results in the prognosis of patients and 5.26% increased chance of favorable prognosis for the antecedent groups respectively. The rule {Duration=Less than 5 days, Caste=Janajati} → {Prognosis=Favorable} (support = 0.3, lift=1.0526) shows that 30% of Janajati patients diagnosed early have a favorable prognosis, with an 5.26% higher likelihood of recovery compared to others. Similarly, {Age Group=18–35, Sex=Male, Duration=Less than 5 days} → {Prognosis=Favorable} (support = 0.2333, lift=1.0526) highlights the importance of early diagnosis in young males. {District=Kathmandu} → {Prognosis=Favorable} (support = 0.2333, lift=1.0526) suggests that patients from Kathmandu have a 5.26% higher chance of favorable recovery, potentially due to better healthcare access. However, none of the association rules for Kala-azar were found statistically significant (P-value > 0.05), though bootstrapping validation showed high support and confidence values, confirming the robustness of the findings (percentage of repetition of rule>70%).

DISCUSSION

The results from this study demonstrate notably lower lift values than previously published research on infectious diseases such as dengue and scrub typhus that employed similar registry data. For instance, Chinedozie (2023),[13] and Rai et al (2023),[14] reported lift values for tuberculosis and COVID-19 prediction which were above 1.1 and 1.5, respectively, despite both studies using datasets of comparable sizes. Nevertheless, the results of this study are consistent with the support and confidence metrics reported by Jahangir et al (2018),[10] suggesting that the results from the association rule mining are stable. One key similarity for all the diseases studied—scrub typhus, dengue, and kala-azar— is the vital need for early diagnosis and intervention, particularly within the first five days of the illness, which corroborates the clinical understanding of acute febrile diseases. The confidence of 1.00 in the rules for scrub typhus and kala-azar further validates the importance of timely intervention for favorable outcomes. These findings underscore the urgent need for improved data collection systems, particularly through digitization, to consolidate public health initiatives for early detection and treatment of diseases.

According to our analysis, patients residing in Dhading district had a probability of 24.6% of obtaining a favorable prognosis. Furthermore, they had 8.33% higher odds of recovering compared to patients from other locations. The same age group of 45-59 years had an 8.33% higher probability of favorable recovery. Other positive prognostic indicators included being female, Brahmin, and having multiple symptoms, which all correspond to findings predicted by Karki et al (2020),[27] regarding scrub typhus prognosis, which underscored the importance of early diagnosis along with certain socio-demographic variables.

Considering dengue cases, patients in the 45-59 age group had a 17.78% probability of a favorable prognosis, with a 5.69% increased probability compared to other age brackets. Additionally, females of the Janajati caste with milder symptoms and fewer overall symptoms had a 6.13% increased recovery probability. Additionally, males who had the disease for 5-10 days also demonstrated improved outcomes. These findings correlate with Prattay et al (2022),[28] which portrays age, gender, and symptom severity as primary prognostic factors in dengue.

For kala-azar, the most positive prognoses were seen in patients diagnosed in the first five days of the illness, especially those between the ages of 18 and 35. Prognoses were also better in males, patients from Kathmandu, and those from the Janajati caste. Some of the association rules related to kala-azar were not statistically significant, likely due to a small sample and the need to explore the issue more deeply.[29]

Among the three illnesses, early diagnosis was the most critical factor. Differences in the socio-demographic factors of age, sex, and caste also influenced the outcomes and these factors interrelated differently between the illnesses. Significant associations were found in both scrub typhus and dengue, while sample size issues likely explain the lack of significant associations in kala-azar. Although there are considerably more dengue cases in the sample, association rule mining is most effective with frequently occurring cases. Some important association rules, which are rare, might not be calculated due to the sample size, but the need for early treatment and individualized approaches to these conditions is clear and critical.

CONCLUSION

Association Rule Mining (ARM) was applied to determine the prognostic patterns of dengue and scrub typhus cases in Kathmandu, Nepal. This research underlines the vital importance of early diagnosis and early treatment in relation to prognosis of scrub typhus, dengue and kala-azar. The results of the study noted the importance of early intervention within five days of the illness. The study further pointed out the impact of socio-demographic factors, namely age, sex and caste, although the impact was relative. The lower lift values, while disappointing, were outweighed by the support and confidence metrics remaining consistent. The findings of this study indicate that although not all association rules with high lift values are statistically significant, their consistent confidence and support levels suggest the reliability and validity of these rules across resampled datasets. More effort is needed to broaden the scope of the analysis, on collection of new data by methods like digitization of the records and strengthening the data collection to broad prospective study. The results of the analysis will assist in targeting early treatment and tailoring intervention strategies to improve the outcomes, for these the study presents new perspectives to practitioners. Although limited by the size of the data, the study demonstrates the prognostic value of ARM in the study of infectious diseases. The limited dataset, nevertheless, is the factor that is limits the scope of findings. Furthermore, expanded electronic health databases would enhance future predictive modeling and improve evidence-based clinical practice, underscoring the need for digitization of hospital records in Nepal.

REFERENCES

1. Lamba D, Hsu WH, Alsadhan M. Predictive analytics and machine learning for medical informatics: A survey of tasks and techniques. 2021 Jan 1 [cited 2025 Jan 12]; Available from: https://www.sciencedirect.com/science/article/pii/B9780128217771000239
2. Rijal KR, Adhikari B, Ghimire B, Dhungel B, Pyakurel UR, Shah P, Bastola A, Lekhak B, Banjara MR, Pandey BD, Parker DM, Ghimire P. Epidemiology of dengue virus infections in Nepal, 2006–2019. Infect Dis Poverty. 2021 Dec;10(1):52.
3. Koirala K. Epidemiology of persistent febrile illnesses in Eastern Nepal [Internet]. Université de Genève; 2021 [cited 2025 Jan 22]. Available from: https://archive-ouverte.unige.ch/unige:150873
4. Abeysinghe R, Cui L. Query-constraint-based mining of association rules for exploratory analysis of clinical datasets in the National Sleep Research Resource. BMC Med Inform Decis Mak. 2018 July;18(S2):58.
5. Ghorbani R, Ghousia R. Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. Int J Data Netw Sci. 2019;3:47–70.
6. Ostapchuk V. Implementation of an interactive pattern mining framework on electronic health record datasets [Internet] [Master Thesis]. [United States]: Heidelberg University; 2019 [cited 2025 Jan 12]. Available from: https://core.ac.uk/download/199187552.pdf
7. Adam A, Garko A. Association rule mining on medical data to predict heart disease. Int J Sci Technol Manag. 2015 Aug 1;4(8):26–35.
8. Domadiya N, Rao UP. Privacy Preserving Distributed Association Rule Mining Approach on Vertically Partitioned Healthcare Data. Procedia Comput Sci. 2019 Jan 1;148:303–12.
9. Sengupta D, Sood M, Vijayvargia P, Hota S, Naik PK. Association rule mining based study for identification of clinical parameters akin to occurrence of brain tumor. Bioinformation. 2013 June 29;9(11):555–9.
10. Jahangir I, Hannan A, Javed S. Prediction of dengue disease through data mining by using modified apriori algorithm. In: Proceedings of the 4th ACM International Conference of Computing for Engineering and Science [Internet]. 2018 [cited 2025 July 30]. p. 1–4. Available from: https://dl.acm.org/doi/abs/10.1145/3213187.3287612
11. Liu Y, Ren L. Application of Association Rule Algorithm in Chronic Disease Prediction. In: 2024 International Conference on Intelligent Computing and Robotics (ICICR) [Internet]. 2024 [cited 2025 July 30]. p. 137–44. Available from: https://ieeexplore.ieee.org/document/10659762
12. Brossette SE, Hymel PA. Data Mining and Infection Control. Clin Lab Med. 2008 Mar 1;28(1):119–26.
13. Chinedozie GC. Emerging patterns of tuberculosis drug-resistant by data mining with association rules (Apriori): using the R-library tbdr19prediction. Federal University of Rio de Janeiro; 2023.
14. Rai VK, Chakraborty S, Chakraborty S. Association rule mining for prediction of COVID-19. Decis Mak Appl Manag Eng. 2023 Apr 8;6(1):365–78.
15. Singh A, Singh D, Upreti K, Sharma V, Rathore BS, Raikwal J. Investigating New Patterns in Symptoms of COVID-19 Patients by Association Rule Mining (ARM). J Mob Multimed. 2022 Aug 25;19(1):1–28.
16. Downs SM, Wallace MY. Mining association rules from a pediatric primary care decision support system. In: Proceedings of the AMIA Symposium [Internet]. 2000 [cited 2025 July 30]. p. 200–4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243862/
17. Sarıyer G, Öcal Taşar C. Highlighting the rules between diagnosis types and laboratory diagnostic tests for patients of an emergency department: Use of association rule mining. Health Informatics J. 2020 June 1;26(2):1177–93.
18. Althammer A, Prückner S, Gehring GC, Lieftüchter V, Trentzsch H, Hoffmann F. Systemic review of age brackets in pediatric emergency medicine literature and the development of a universal age classification for pediatric emergency patients - the Munich Age Classification System (MACS). BMC Emerg Med. 2023 July 25;23:77.
19. Bland JM, Altman D. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. The Lancet. 1986 Feb 8;327(8476):307–10.
20. Sadler K, Alwali M, Nancarrow L, Albalawi M, Khan S, Alyami HH. Discharge against medical advice in pediatrics: A review and recommendation of a five-component approach. Int J Pediatr Adolesc Med. 2023 Sept;10(3):51.
21. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. ACM SIGMOD Rec. 1993 June;22(2):207–16.
22. Shmueli G, Bruce PC, Yahav I, Patel NR, Jr KCL. Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. John Wiley & Sons; 2018. 576 p.
23. Fong J, Huang SM, Hsueh HY. Online analytical mining association rules using Chi-square test. Int J Bus Intell Data Min. 2007;2(3):311.
24. Jaiswal V, Agarwal J. The Evolution of the Association Rules. Int J Model Optim. 2012;726–9.
25. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. 1st ed. New York: Chapman and Hall/CRC; 1994. 456 p.

26. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1995. p. 1137–43. (IJCAI'95).
27. Karki K, Acharya B, Dhimal M, Aryal K, Sharma K, Khanal P, Ranabhat K, Dumre S, Jha A. Descriptive Epidemiology of Scrub Typhus in Nepal, 2017. Kathmandu: Nepal Health Research Council; 2017.
28. Prattay KMR, Sarkar MR, Shafiullah AZM, Islam MS, Raihan SZ, Sharmin N. A retrospective study on the socio-demographic factors and clinical parameters of dengue disease and their effects on the clinical course and recovery of the patients in a tertiary care hospital of Bangladesh. PLoS Negl Trop Dis. 2022 Apr 4;16(4):e0010297.
29. Mandal A, Chakraborty T, Gurevych I. Towards Privacy-aware Mental Health AI Models: Advances, Challenges, and Opportunities [Internet]. arXiv; 2025 [cited 2025 Sept 25]. Available from: http://arxiv.org/abs/2502.00451.