Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 30s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Exploring Co-occurrence Patterns in Infectious Diseases using Association Rule Mining: A Hospital-based Study from the Cities of Nepal

S. C. Kafle[1*], M. Nain[2], B. Sharma[3], H. P. Adhikari[4], R. Raman[5]

[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]Lecturer, Department of Health Education, Mahendra Multiple Campus, Banke, Nepal

[4]Suvekchya International Hospital, Kathmandu, Nepal

[5]Nita Polyclinic and Diagnostic Centre, Kathmandu, Nepal

Corresponding Author:

S. C. Kafle,

PhD Scholar, Department of Statistics, Om Sterling Global University, Hisar.

Email ID : sarad.kafle49@gmail.com

Cite this paper as: S. C. Kafle, M. Nain, B. Sharma, H. P. Adhikari, R. Raman, (2025) Exploring Co-occurrence Patterns in Infectious Diseases using Association Rule Mining: A Hospital-based Study from the Cities of Nepal. *Journal of Neonatal Surgery*, 14 (30s), 1145-1154

ABSTRACT

The co-occurrence of patient characteristics with infectious diseases provides essential insights into disease distribution, supporting improved preparedness and response strategies in both endemic and epidemic contexts. While Association Rule Mining (ARM) is widely applied in computer science for pattern discovery, its use in healthcare, particularly for understanding demographic and geographic clustering of infectious diseases, remains limited. This study applied the Apriori algorithm, a commonly used ARM technique, to retrospective inpatient data from three tertiary hospitals in Nepal—Kathmandu, Bharatpur, and Pokhara—totaling 1,019 records. Rules were evaluated based on support, confidence, and lift to identify strong, non-redundant associations. The findings showed that Dengue consistently clustered among younger urban populations, with gender and caste (particularly males and Brahmins) emerging as key relationships, and showed distinct geographical concentrations in Kathmandu and Pokhara. In contrast, Scrub typhus was strongly associated with older adults, females, and rural residents, with exceptionally high lift values observed in Chitwan, indicating a substantially elevated risk compared to the general inpatient population. No significant rules emerged for Kala-azar due to its low case frequency. These findings underscore the heterogeneity of vector-borne disease distribution across ecological and demographic contexts. The study concludes that ARM is a valuable analytic tool for uncovering non-random, epidemiologically meaningful patterns, offering practical guidance for targeted interventions, resource allocation, and the development of context-specific public health strategies in Nepal and similar regions.

Keyword: Apriori Algorithm, Frequent mining, Dengue, Scrub typhus

1. INTRODUCTION

Infectious diseases still represent the most significant global public health issue. While COVID-19 emerged in 2019, the proportion of deaths due to infectious diseases jumped to 23% in 2020, surpassing the previous peak of 2012, and then rose again to 28.1% in 2021, a figure comparable to that in 2005 (WHO, 2024). This global death toll is primarily due to the COVID-19 pandemic during this period, and the recurring occurrence of such endemic and pandemic infectious diseases has prompted scientists and governments around the world to study the patterns and co-occurrence of these diseases in detail. Meanwhile, Nepal was not far from the problems faced by other countries in the contemporary situation. However, there is a significant decrease in the burden of infectious diseases, from 70% to 21.8%, between 2000 and 2021 (Mishra et al., 2025). In Nepal, particularly in urban areas, a dengue outbreak has been reported. Kathmandu, Pokhara, Bharatpur, Biratnagar, and Dharan were affected by this disease in previous years. The data of the Epidemiology and Disease Control Division in Teku, Kathmandu, Nepal, in 2023 showed that Kaski, Tanahun, Chitwan, and Bara districts are among the top ten reporting dengue cases and are also potential hotspots for other infectious diseases (EDCD, 2023). These districts have dense urban populations, underscoring the need to study infectious diseases such as dengue to better understand their co-occurrence. Additionally, people infected with dengue are also at risk of contracting scrub typhus, particularly those from the outskirts (Jose et al., 2022). Although healthcare in Nepal has improved, this may be due to rising hospital attendance driven by advances in health technology and increased healthcare infrastructure. However, the exposure risk to infectious agents, even.

within the hospital environment, has also increased. However, the frequent waves of pandemic and endemic diseases have startled concerned bodies, prompting them to better understand these diseases. This highlights the need for enhanced knowledge and control of infectious disease patterns, particularly those that coexist within hospitals (Baker et al., 2022)

These recent COVID-19, SARS, MERS, and Ebola outbreaks have reinforced the value of monitoring and investigating patterns of infectious diseases (Han et al., 2023). Bacterial and viral infections account for the most prevalent infectious diseases in Nepal, often originating from zoonotic sources and sustained through human-to-human transmission (Gautam et al., 2021). Even with attempts to control diseases using molecular methods and bioinformatics-based information tools, the limited application of data mining techniques, specifically association rule mining (ARM), has been used to identify disease co-occurrence in Nepal. However, this technique has already been used for the prediction of diseases based on the history and symptoms of patients for dengue (Gómez-Pulido et al., 2020; Jahangir et al., 2018; Shakil et al., 2015), scrub typhus, Tuberculosis (Chinedozie, 2023), and COVID-19 (Rai et al., 2023; Singh et al., 2022). It is a raging pandemic that has created havoc with its impact ranging from loss of millions of human lives to social and economic disruptions of the entire world.

The limited use of ARM for disease prediction shows the research gap and makes this research meaningful. Association rule mining is an unsupervised data mining process for discovering significant “If-Then” patterns from large datasets. Commonly used in market basket analysis, healthcare, and finance, ARM provides a scalable, straightforward solution for identifying hidden relationships in complex datasets (Shmueli et al., 2018). Popular algorithms such as Apriori, FP-Growth, and Eclat attest to its efficacy in finding rules specified by support, confidence, and lift (Rai et al., 2023). In the clinical context, ARM can discover co-occurrence patterns among diseases, facilitate the identification of hidden interactions, and support early diagnosis, clinical decisions, and treatment plans. Although existing research has demonstrated the applicability of ARM to hospital data (Rai et al., 2023; Singh et al., 2022), there has been limited research on infectious disease co-occurrence in Nepal. Co-occurrence patterns, also known as the frequent concurrent appearance of patients with a particular disease, are observed. For instance, a specific group of people might have a higher chance of exposure to a particular infectious disease. In Nepal, people from the outskirts and lower-income groups face a higher risk of vector-borne diseases, such as scrub typhus. It is necessary to investigate which age groups, genders, caste groups, and localities have high exposure to such infectious diseases; this is the study of co-occurrence. The co-occurrence patterns also facilitate the development of specific clinical guidelines and systems for detecting early outbreaks.

The objective of this study is to identify frequent demographic patterns associated with specific infectious diseases using association rule mining. This objective aims to close the current research gap by exploring the co-occurrence patterns of communicable diseases in hospitals in Nepal using ARM techniques. Through the extraction of association rules from registry data, the study aims to uncover hidden associations between infectious diseases and contribute to the development of more effective public health policies and patient care. The results of the survey shall be the identification of essential disease combinations and comprehension of the diagnostic and epidemiologic influence of the associations.

2. METHODOLOGY

This study is based on a retrospective analysis of one-year registry data from inpatients admitted to a selected hospital between January 1, 2024, and December 30, 2024. The selected hospitals for this study are Sukraraj Tropical and Infectious Disease Hospital (STIDH), Teku, Kathmandu; Bharatpur Hospital (BH), Bharatpur, Chitwan; and Western Regional Hospital, Pokhara (WRH), Kaski. These hospitals were selected to reflect the spread of dengue disease across the top five districts of Nepal. According to the 2023 health report of the Epidemiology and Disease Control Division in Teku, Kathmandu, Nepal, Kaski, Tanahun, Chitwan, and Kathmandu districts are among the top 10 districts reporting the most dengue cases (EDCD, 2023). To include patients from these districts, the three aforementioned hospitals were selected, as residents of the district and neighboring areas typically go there for medication. For this study, a retrospective method is chosen for data collection because it would take a year if conducted using a longitudinal approach (Blane, 1996). Before data collection, a consent letter from the respective hospitals is obtained and submitted to the Nepal Health Research Council for ethical approval. After receiving NHRC approval, data is collected in accordance with the hospital's rules and regulations. Registry data is collected from the discharge books of the respective hospitals and entered into an Excel file, then analyzed using ARM in the RStudio platform. The inclusion rule is that the patient must be an inpatient at the selected hospital during the study period and have dengue and scrub typhus. The other patients were excluded from further analysis. Three hundred fifty-four inpatients from STIDH, four hundred sixty from BH, and two hundred five inpatients from WRH were included in the data analysis.

This study used Apriori Approach of ARM proposed by Agrawal et al. (1993) (Agrawal et al., 1993), this rule starts with the generation of frequent itemsets with just one item and then recursively generates frequent itemsets with two items, then with three items, and so on, until frequent itemsets of all sizes have been generated (Shmueli et al., 2018). The strength of different association rules derived during the analysis was compared using support, confidence, and lift ratio (Shmueli et al., 2018).

Based on the antecedent and consequent, an association rule is derived. Antecedent is used to describe the IF part, and resultant is used to describe the THEN part of the co-occurrence of itemsets (Shmueli et al., 2018). Confidence is the ratio of the number of transactions that include all antecedent and consequent itemsets to the number of transactions that include all the antecedent itemsets. They are called the support of the association rule.

In this study, the strength of association rules derived through data mining was evaluated using support, confidence, and lift, three fundamental metrics. While confidence represents the conditional probability of the consequent given the antecedent and indicates how often the rule holds, it can be misleadingly high when both antecedent and consequent are individually frequent, even if they occur independently. To address this, a lift was calculated, which quantifies how much more often the antecedent and consequent occur together than would be expected under independence. A lift value greater than 1 indicates a meaningful positive association. The greater the lift, the more substantial and more statistically significant the relationship between antecedent and consequent (Shmueli et al., 2018). By combining these metrics, this methodology ensures that only rules with both high reliability and substantive association are used, enhancing the validity of the patterns extracted from patient data.

3. RESULTS

The demographic distribution of inpatients across age groups shows that STIDH recorded the highest proportion of patients aged 19–45 years (68.08%), significantly surpassing the proportions at BH (43.7%) and WRH (42.93%), indicating a predominance of young and middle-aged adults at this national referral center. Conversely, infants were absent in both BH and STIDH, while WRH had the highest representation at 3.41%, highlighting its broader age-based service outreach. The over-65 age group was most prevalent in BH (25.65%), significantly exceeding WRH (5.37%) and STIDH (4.8%), suggesting that BH may cater more to geriatric cases. In the under-18 category, WRH again had the highest share (30.73%), contrasting sharply with the lowest in BH (3.48%). Regarding gender, BH reported the highest proportion of female inpatients at 64.78%, suggesting possible gender-related service preference or accessibility, whereas STIDH had the lowest female representation (37.57%), with a male predominance (62.43%). In contrast, WRH displayed a relatively balanced gender distribution. Regarding the residential location of patients, divided into urban and rural regions, urban residents predominated at STIDH (71.75%), followed by WRH (66.83%) and BH (53.7%), reflecting STIDH's location in the capital and its role as a specialized tertiary hospital. Conversely, BH recorded the highest percentage of rural inpatients (46.3%), while STIDH had the lowest rural representation (28.25%), indicating differing catchment areas and accessibility across the institutions (see Table 1).

Table 1: Demographic Characteristics of Inpatients in Three Hospitals

	BH	WRH	STIDH	Total
Age group				
Infants	0(0%)	7(3.41%)	0(0%)	7(0.69%)
19-45	201(43.7%)	88(42.93%)	241(68.08%)	530(52.01%)
46-65	125(27.17%)	36(17.56%)	72(20.34%)	233(22.87%)
Above 65	118(25.65%)	11(5.37%)	17(4.8%)	146(14.33%)
Below 18	16(3.48%)	63(30.73%)	24(6.78%)	103(10.11%)
Total	460(100%)	205(100%)	354(100%)	1019(100%)
Gender				
Female	298(64.78%)	92(44.88%)	133(37.57%)	523(51.32%)
Male	162(35.22%)	113(55.12%)	221(62.43%)	496(48.68%)
Total	460(100%)	205(100%)	354(100%)	1019(100%)
Locations				
Rural	213(46.3%)	68(33.17%)	100(28.25%)	381(37.39%)
Urban	247(53.7%)	137(66.83%)	254(71.75%)	638(62.61%)
Total	460(100%)	205(100%)	354(100%)	1019(100%)

Patients' distribution of BH which is located in Chitwan shows that it primarily serves patients from Chitwan district, where it is situated, which accounts for a dominant 63.91% of its inpatients. Other notable contributors include Nawalpur (10%), Gorkha (8.26%), and Tanahun (6.96%), highlighting a strong regional focus within the neighboring district of Chitwan, located in Bagmati Province. At the same time, WRH Pokhara exhibits a similar pattern, with 59.02% of patients coming from Kaski District, the district where the hospital is situated. It also draws significantly from Tanahun (18.05%) and Syangja (13.17%), reflecting its service to the wider Gandaki Province. Meanwhile, Sukraraj Hospital in Kathmandu has a more diverse patient demographic, though Kathmandu District still accounts for the majority at 51.98%. Patients also originate from Lalitpur (12.15%), Dhading (8.76%), Sarlahi (2.54%), and Nawalparasi (1.95%), supporting its role as a national referral center attracting patients from multiple districts beyond the valley (see Figure 1).

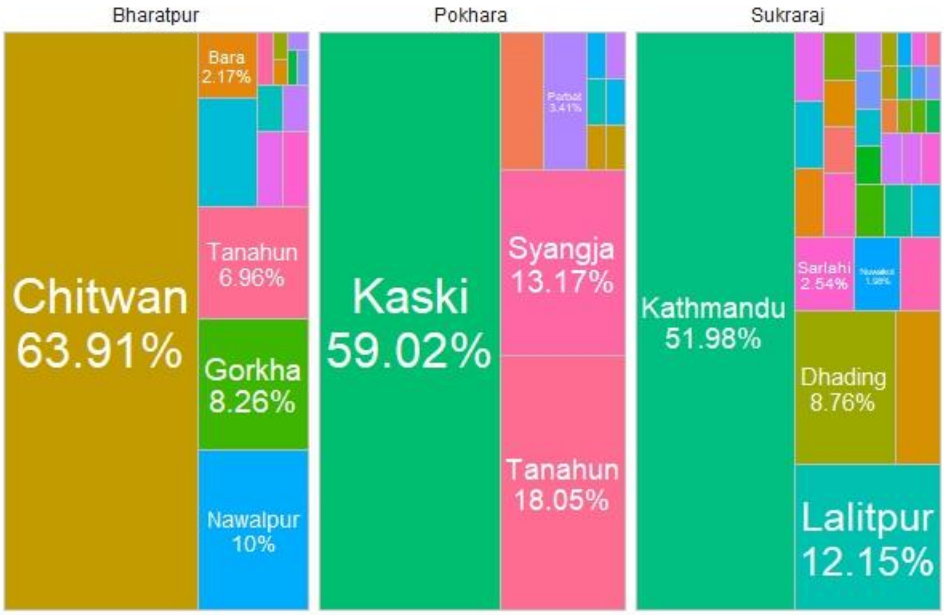

Figure 1: Distribution of Patients by their district

Discussing the local address of patients, WRH Pokhara had the highest concentration of patients from a single area, with 52.2% residing in Pokhara, indicating a predominantly local catchment. Bharatpur Hospital followed, with 44.13% of its inpatients from Bharatpur, while also serving nearby municipalities such as Gaindakot (6.09%) and Khairahani (5.22%). In contrast, Sukraraj Tropical and Infectious Disease Hospital Kathmandu, displayed the most diverse patient demographic, with the largest share from Sanepa (4.8%), followed by Kalimatī (4.24%), Dhadhing (3.95%), and Kalanki (3.39%), highlighting its role as a national referral center drawing patients from a broader geographical area (see Figure 2).

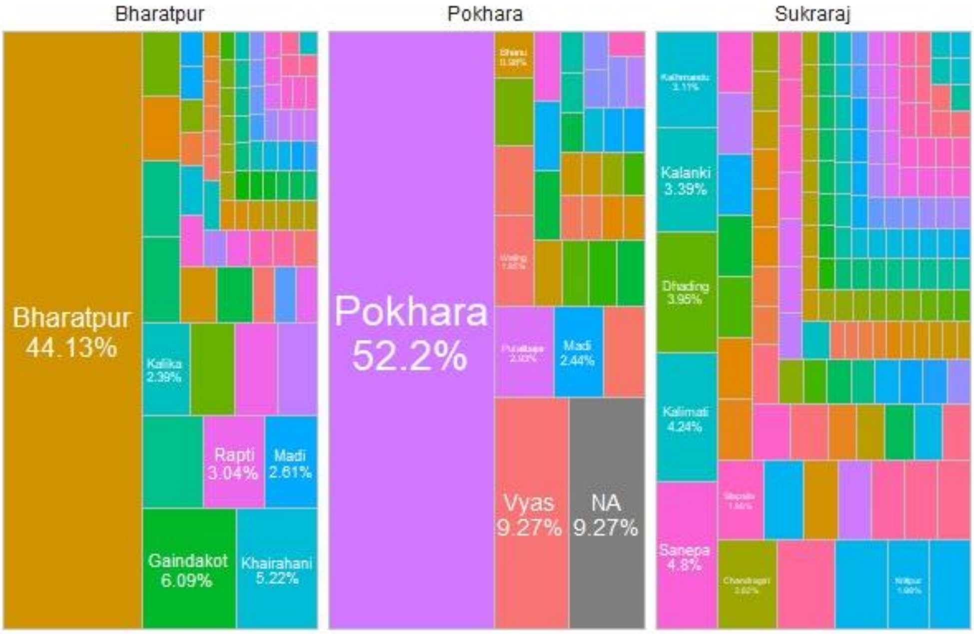

Figure 2: Distribution of Patients by their Local Level Address

While analyzing the distribution of inpatients by caste, Brahmin patients made up the largest group at all three hospitals, most notably in WRH (39.51%), followed by BH (33.7%), and STIDH (30.23%), reflecting their significant presence in urban and educated populations. Scheduled Caste patients were notably higher at WRH (22.44%) than at BH (12.39%) and STIDH (6.5%), indicating greater use of public health services by marginalized groups in the Pokhara region. At STIH, Newar patients comprised 12.99%, the highest among the three, likely due to the hospital's location in the Kathmandu Valley, where Newars are indigenous. Interestingly, the Tamang, Kshetri, and Magar groups each accounted for around 10% at STIDH, suggesting a more balanced ethnic representation. In Bharatpur, Gurung (8.91%) and Magar (8.48%) had slightly higher representation than in WRH and STIDH, while Chepang (4.78%), a marginalized indigenous group, also appeared to be noticeably represented (see Figure 3).

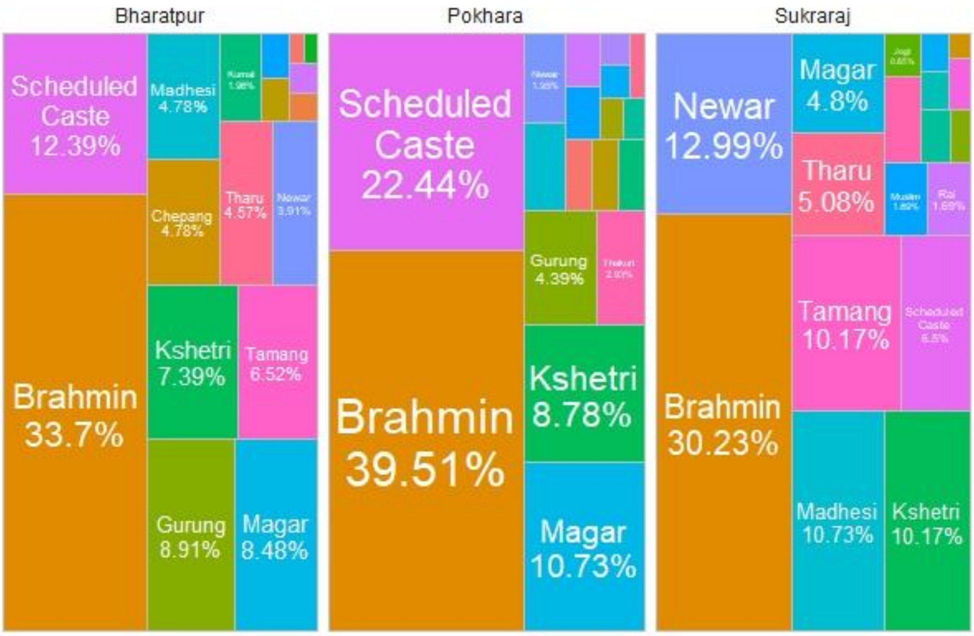

Figure 3: Distribution of Patients by their Caste

The Apriori algorithm generated 688 rules for STIDH patients, out of which eight rules with support greater than 0.10 and lift values above 1.10 were selected for detailed interpretation. These selected rules, summarized in Table 2, highlight meaningful associations between demographic subgroups and Dengue diagnosis.

The first rule highlights that the subgroup of females aged 19–45 years residing in urban Kathmandu shows a support of 11.58%, indicating that this group constitutes about 11.6% of the entire patient population. A confidence level of 1.0 suggests that every patient in this subgroup was diagnosed with Dengue. The lift of 1.19 indicates that Dengue occurrence in this group is 19% more likely than expected under independence, highlighting an influential association.

The second rule exhibits that Patients from urban areas of Lalitpur district have a support of 10.17% and a confidence of 0.973, meaning that 97.3% of patients in this subgroup were diagnosed with Dengue. The lift value of 1.16 indicates that urban Lalitpur patients were 16% more likely to present with dengue than the baseline prevalence at STIDH. This demonstrates that urban residence in Lalitpur is a relevant contextual factor for Dengue diagnosis.

The third rule explains that the subgroup of males aged 19–45 years residing in urban Kathmandu constitutes 17.51% of the dataset (support) with a confidence of 0.969. This means that nearly 97% of patients in this subgroup had Dengue. The lift of 1.16 indicates that Dengue was 16% more likely in this subgroup than expected by chance, confirming that age, gender, and urban residence are consistent predictors of Dengue.

The fourth rule explains that patients aged 19–45 years from the Brahmin caste residing in urban Kathmandu have a support of 11.02% and a confidence of 0.951, showing that more than 95% of patients in this cluster were diagnosed with Dengue. The lift of 1.14 indicates a 14% higher likelihood compared to baseline prevalence. This rule further suggests that caste, alongside age and residence, may contribute to clustering of Dengue cases in the study population.

The fifth rule is that the broader category of urban male patients from any district has the largest subgroup, with a coverage of 45.48% of the total patient population and a support of 42.09%. The confidence of 0.926 shows that more than 92% of patients in this group were diagnosed with Dengue. With a lift of 1.11, this subgroup was 11% more likely to have Dengue than expected under independence. While broader, this rule complements Rule 3 and demonstrates the influence of gender and urban residence at a general level.

The sixth rule, based on the previous rule, indicates that urban Brahmin patients aged 19–45 years show a support of 15.25% and a confidence level of 0.931. This subgroup was 11% more likely to be diagnosed with Dengue (lift = 1.11) compared to baseline, with a coverage of 16.38% of the patient population. This rule refines the broader male–urban association by introducing caste, highlighting an additional dimension of clustering.

Finally, the seventh rule for the subgroup of females aged 19–45 years, which represents 23.73% of the dataset, shows that 92.3% of patients in this cluster were diagnosed with Dengue. The lift value of 1.10 indicates that Dengue was 10% more likely in this subgroup than expected by chance. This finding reinforces the vulnerability of young adult females to Dengue, consistent with Rule 1, and demonstrates that gender and age remain strong determinants even without specifying location.

Taken together, these rules highlight the disproportionate clustering of Dengue cases among young adults (19–45 years), urban residents, and within specific caste (Brahmin) groups. While confidence values across rules are consistently high, the lift values, ranging from 1.10 to 1.19, confirm that these associations are stronger than expected under independence and therefore meaningful. Across the selected rules, three consistent factors emerge: age (19–45 years), urban residence, and gender. These variables appear repeatedly in high-confidence and high-lift rules, suggesting that young adults living in urban environments are disproportionately vulnerable to Dengue. Gender-specific clustering is also evident, with both male and female subgroups exhibiting high Dengue prevalence; however, females aged 19–45 years show slightly stronger associations. The inclusion of caste, specifically Brahmin, in two of seven rules indicates an additional layer of social or behavioral determinants influencing Dengue diagnosis. Collectively, these findings underscore that Dengue is not evenly distributed across the patient population but instead follows identifiable demographic and contextual patterns.

Table 2: Association rules using the Apriori Algorithm for STIDH patients

SN	Antecedent	Consequent	Support	Confidence	coverage	Lift
1	19 to 45 years, Female, Kathmandu, Urban	Dengue	0.1158	1	0.1158	1.19
2	Lalitpur, Urban	Dengue	0.1017	0.973	0.1045	1.16
3	19 to 45 years, Male, Kathmandu, Urban	Dengue	0.1751	0.9688	0.1808	1.16
4	19 to 45 years, Brahmin, Kathmandu, Urban	Dengue	0.1102	0.9512	0.1158	1.14
5	Male, Urban	Dengue	0.4209	0.9255	0.4548	1.11
6	19 to 45 years, Brahmin, Urban	Dengue	0.1525	0.931	0.1638	1.11

7	19 to 45 years, Female	Dengue	0.2373	0.9231	0.2571	1.10

The Apriori algorithm generated twenty-three association rules for BH patients. After filtering based on support greater than 0.10 and lift above 1.10, seven rules were identified as meaningful and are presented in Table 3. These rules highlight the demographic and geographic patterns underlying Dengue and Scrub typhus diagnoses among the BH patient Population.

The first association rule is described for the subgroup of patients aged 19–45 years from urban Bharatpur, Chitwan, which accounts for 16.5% of the total dataset (support). Within this group, 80% were diagnosed with Dengue (confidence = 0.80). The lift value of 1.78 indicates that the occurrence of Dengue in this subgroup is 78% more likely than expected under independence, making this the strongest Dengue-related association observed for BH.

The second rule states that patients aged 65 years or older who are female represent 10.9% of the dataset. Among them, 70.4% were diagnosed with Scrub typhus. The lift value of 1.73 suggests that elderly females are 73% more likely to be diagnosed with Scrub typhus than the baseline prevalence would suggest, pointing to a strong age–gender association.

The third rule describes the subgroup of females aged 19–45 years residing in urban Bharatpur, which accounts for 10.2% of the dataset and 77.1% of those diagnosed with Dengue. The lift value of 1.71 shows that the likelihood of Dengue among this subgroup is 71% higher than expected under independence.

The fourth rule, closely related to the third rule, shows a subgroup of females aged 19–45 years residing in urban Bharatpur, Chitwan, with a support of 10.2%, confidence of 0.771, and a lift of 1.71. This rule refines the subgroup by explicitly including Chitwan, but the interpretation remains consistent: young adult urban females in Bharatpur face disproportionately higher risk of Dengue.

The fifth rule encompasses a broader category of patients aged 65 years and above, comprising 16.7% of the dataset, of whom 65.3% were diagnosed with Scrub typhus. The lift of 1.61 confirms that elderly patients are 61% more likely than average to present with Scrub typhus, reinforcing the significance of age as a determinant.

The sixth rule narrows down rule five; the subgroup of patients aged 65 and above from Chitwan constitutes 10.2% of the dataset. Within this group, 62.7% were diagnosed with Scrub typhus, with a lift of 1.54. This suggests that being both elderly and residing in Chitwan increases the relative likelihood of Scrub typhus diagnosis.

Finally, rule seven explains the subgroup of patients residing in rural areas of Chitwan, with a support of 11.1% and a confidence of 0.622, indicating that nearly two-thirds of these patients had Scrub typhus. The lift value of 1.53 indicates a 53% higher likelihood of Scrub typhus than expected, highlighting rural residence as an additional contextual factor.

The rules derived for the patients admitted to BH reveal two distinct patterns of disease clustering. First, Dengue is strongly associated with younger adults (19–45 years), particularly females, residing in urban Bharatpur, Chitwan. These groups show lifts of 1.71-1.78, reflecting a substantially elevated risk. Second, Scrub typhus is concentrated among older patients (age 65+), with an additional risk among females and those living in rural Chitwan. Collectively, these findings suggest that while Dengue disproportionately affects young urban populations, Scrub typhus clusters among elderly and rural subgroups, underscoring the need for disease-specific prevention and intervention strategies tailored to demographic and geographic risk factors (see Table 3).

Table 3: Association rule based on the Apriori algorithm for the patients of BH

SN	Antecedent	Consequent	Support	Confidence	Coverage	Lift
1	19 to 45 years, Chitwan, Bharatpur, Urban	Dengue	0.1652	0.8	0.2065	1.78
2	Above 65 years, Female	Scrub	0.1087	0.7042	0.1544	1.73
3	19 to 45 years, Female, Bharatpur, Urban	Dengue	0.1022	0.7705	0.1326	1.71
4	19 to 45 years, Female, Chitwan, Bharatpur, Urban	Dengue	0.1022	0.7705	0.1326	1.71
5	Above 65 years	Scrub	0.1674	0.6525	0.2565	1.61
6	Above 65 years, Chitwan	Scrub	0.1022	0.6267	0.1630	1.54
7	Chitwan, Rural	Scrub	0.11087	0.621951	0.178261	1.53

Application of the Apriori algorithm to WRH patient data yielded 526 rules. After applying thresholds of support above 0.10 and lift above 1.10, ten rules were selected as particularly meaningful. These rules, presented in Table 4, highlight demographic and locational factors associated with Dengue diagnosis among WRH patients.

The first rule describes the subgroup of male patients under 18 years from the urban area of Pokhara metropolitan city in

Kaski district, which represents 10.2% of the dataset. Within this group, 91.3% were diagnosed with Dengue (confidence = 0.913). The lift of 1.25 indicates that this subgroup is 25% more likely to have Dengue compared to the baseline prevalence, suggesting strong clustering among young urban males.

The second rule explains that the patients belonging to the Brahmin caste residing in urban Pokhara, Kaski, make up 19.5% of the dataset, with 88.9% of them diagnosed with Dengue. The lift of 1.21 indicates a 21% higher likelihood of Dengue than expected under independence, showing that caste and residence are relevant contributors to Dengue clustering.

The third rule applies to the subgroup of Brahmin patients aged 19–45 years living in urban areas, which accounts for 10.7% of the dataset, and 88% of whom are diagnosed with Dengue. The lift of 1.20 indicates a 20% higher likelihood of Dengue than expected. This finding reinforces the importance of age and caste in shaping Dengue vulnerability.

The fourth rule states that the cluster of male patients below 18 years represents 17.6% of the dataset. Among them, 87.8% were diagnosed with Dengue. With a lift of 1.20, this subgroup shows a 20% higher likelihood of Dengue compared to the baseline, underscoring the susceptibility of children and adolescents, particularly boys.

The fifth rule, Which Narrows down the fourth rule, indicates that male patients under 18 years of age residing in Pokhara show a support of 10.2% and a confidence level of 0.875. This subgroup is 20% more likely to have Dengue than expected under independence (lift = 1.20). This emphasizes that location strengthens the association among young males.

The sixth rule states that the subgroup of Brahmin patients aged 19–45 years residing in Kaski accounts for 10.2% of the dataset, with 87.5% diagnosed with Dengue. The lift value of 1.20 indicates a 20% higher likelihood of Dengue than in the general patient population, consistent with Rule 3.

The seventh rule highlights that patients aged 19–45 years who are male and reside in Pokhara form another subgroup with 10.2% support, 87.5% confidence, and a lift of 1.20. This rule highlights the clustering of Dengue among adult males in urban centers.

The eighth rule states that male patients under 18 years of age residing in urban Pokhara show a support of 10.2% and a confidence of 0.875. The lift of 1.20 again demonstrates that young urban males are consistently more likely to present with Dengue.

The ninth rule applies to males aged 19–45 years from urban Pokhara, Kaski, who form 10.2% of the dataset, with 87.5% diagnosed with Dengue. The lift of 1.20 suggests that Dengue prevalence in this group is 20% higher than baseline, reinforcing the gender–age–urban link.

Finally, the tenth rule explains that male Brahmin patients from urban Pokhara, Kaski, represent 12.7% of the dataset, with 86.7% of them diagnosed with Dengue. The lift value of 1.18 indicates that Dengue is 18% more likely in this subgroup, reflecting the combined influence of gender, caste, and location.

The WRH findings reveal a consistent pattern: Dengue clustering is strongly associated with younger age (below 18 or 19–45 years), male gender, and urban residence in Pokhara, Kaski. Several rules highlight the heightened risk among boys and young men, while some of the additional contribution of caste (Brahmin) to Dengue prevalence is also found interesting. The lift values across rules (1.18–1.25) confirm that these associations are meaningfully stronger than expected by chance. Taken together, the results suggest that Dengue risk at WRH is concentrated among younger, urban, and caste-specific (Brahmin) male subgroups, pointing to demographic and sociocultural dimensions that shape disease vulnerability (see Table 4).

Table 4: Association Rules using Apriori for the Patients of WRS

SN	Antecedent	Consequent	Support	Confidence	Coverage	Lift
1	Below 18, Male, Kaski, Pokhara Urban	Dengue	0.1024	0.913	0.1122	1.25
2	Brahmin, Kaski, Pokhara, Urban	Dengue	0.1951	0.8889	0.2195	1.21
3	19 to 45 years, Brahmin, Urban	Dengue	0.1073	0.88	0.122	1.2
4	Below 18 years, Male	Dengue	0.1756	0.878	0.2	1.2
5	Below 18 years, Male, Pokhara	Dengue	0.1024	0.875	0.1171	1.2
6	19-45 years, Brahmin, Kaski	Dengue	0.1024	0.875	0.1171	1.2
7	19-45 years, Male, Pokhara	Dengue	0.1024	0.875	0.1171	1.2
8	Below 18 years, Male, Pokhara, Urban	Dengue	0.1024	0.875	0.1171	1.2
9	19 to 45 years, Male, Kaski, Pokhara, Urban	Dengue	0.1024	0.875	0.1171	1.2
10	Male, Brahmin, Kaski, Pokhara, Urban	Dengue	0.1268	0.8667	0.1463	1.18

4. DISCUSSION

Although the current study is based on ARM, and this rule is based on frequent pattern discovery, the generated rules reflect associations from high-frequency subgroups in the dataset, meaning that less frequent itemsets may not be adequately represented. This limitation is essential, as it restricts the ability to capture rare disease–demographic combinations. Nonetheless, the selected rules provide meaningful insights into how patients' demographic characteristics and place of residence are associated with diagnoses of Dengue and Scrub typhus. While the study dataset included Dengue, Scrub typhus, and Kala-azar, the relatively low frequency of Kala-azar cases prevented the generation of association rules for this disease. In contrast, Dengue and Scrub typhus showed multiple associations that varied in nature across hospitals, reflecting differences in demographic clustering and geographical context.

Starting with the similar association rule found across the hospitals, in the context of STIDH, Dengue was strongly clustered among urban residents aged 19–45 years, with gender (both male and female subgroups) and caste (Brahmin) further refining these associations. The consistent presence of urban settings indicates that Dengue is essentially an urban-centered disease in Kathmandu.

While in BH, two distinct patterns emerged. Dengue was concentrated among younger, urban populations (particularly females aged 19–45 years in Bharatpur), while Scrub typhus was clustered among elderly patients, especially women and rural residents. This demonstrates how two vector-borne diseases coexist in the same geographic area but affect very different population subgroups.

Similarly, in WRH, Dengue associations were primarily linked to male gender, younger age (below 18 and 19–45 years), urban residence, and Brahmin caste. Unlike BH, no Scrub typhus-related rules were prominent, reflecting the predominance of Dengue in this hospital setting.

The above-mentioned findings highlight that Dengue consistently clusters in urban areas and among younger populations, and that gender and caste also influence transmission. These findings are similar to those of Acharya et al. (2018) and Acharya et al. (2016), which suggest that dengue is higher in urban areas than in rural areas. Whereas Scrub typhus is more prevalent in rural areas and among females, this finding aligns with Gautam et al. (2019), who found that older adults and rural residents are at high risk of exposure to Scrub typhus. However, current research findings contradict the findings of Gautam et al. (2019), who found that females have lower exposure to scrub typhus than males. While it shares similarities with the findings of Linsuwanon et al. (2024), this may be due to differences in disease distribution across geographical locations, as the study by Linsuwanon et al. (2024) was conducted in Chitwan and Kaski, where the current research was undertaken.

5. CONCLUSION

This study demonstrates the potential of association rule mining (ARM) as an alternative analytic approach for uncovering hidden patterns in hospital-based infectious disease data. Despite its inherent limitation of prioritizing frequent patterns over rarer itemsets, ARM revealed distinct and consistent associations between sociodemographic characteristics, place of residence, and disease occurrence. Dengue was found to cluster predominantly among younger, urban populations-often stratified by gender and caste. While Scrub typhus exhibited stronger associations with elderly, rural, and female subgroups, there was notable variability across hospitals. These findings underscore the heterogeneity of disease distribution within Nepal, reflecting both ecological and sociodemographic contexts. By highlighting population subgroups at elevated risk, this study not only aligns with prior epidemiological evidence but also provides a data-driven foundation for designing geographically and socially tailored intervention strategies against vector-borne diseases.

6. ACKNOWLEDGEMENT

We extend our sincere gratitude to all who supported this research. In particular, we are deeply thankful to the administrations of Sukraraj Tropical and Infectious Disease Hospital, Teku, Kathmandu; Bharatpur Hospital, Chitwan; and Western Regional Hospital, Pokhara, Kaski, for their invaluable cooperation and assistance in facilitating data collection

REFERENCES

[1] Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases. ACM SIGMOD Rec. 22, 207–216. https://doi.org/10.1145/170036.170072

[2] Baker, R.E., Mahmud, A.S., Miller, I.F., Rajeev, M., Rasambainarivo, F., Rice, B.L., Takahashi, S., Tatem, A.J., Wagner, C.E., Wang, L.-F., Wesolowski, A., Metcalf, C.J.E., 2022. Infectious disease in an era of global change. Nat. Rev. Microbiol. 20, 193–205. https://doi.org/10.1038/s41579-021-00639-z

[3] Blane, D.B., 1996. Collecting retrospective data: Development of a reliable method and a pilot study of its use. Soc. Sci. Med. 42, 751–757. https://doi.org/10.1016/0277-9536(95)00340-1

[4] Chinedozie, G.C., 2023. Emerging patterns of tuberculosis drug resistance by data mining with association rules

(Apriori): using the R-library tbdr19prediction. Federal University of Rio de Janeiro.

[5] EDCD, 2023. Early Warning and Reporting System (EWARS) annual report 2022 & 2023. Epidemiology and Disease Control Division.

[6] Gautam, A., Upadhayay, P., Ghimre, D., Khanal, A., Gaire, A., Kaphle, K., 2021. Prioritised Zoonotic Diseases of Nepal: A Review. Int. J. Appl. Sci. Biotechnol. 9, 1–15. https://doi.org/10.3126/ijasbt.v9i1.34967

[7] Gautam, R., Parajuli, K., Sherchand, J.B., 2019. Epidemiology, Risk Factors and Seasonal Variation of Scrub Typhus Fever in Central Nepal. Trop. Med. Infect. Dis. 4, 27. https://doi.org/10.3390/tropicalmed4010027

[8] Gómez-Pulido, J.A., Romero-Muelas, J.M., Gómez-Pulido, J.M., Castillo Sequera, J.L., Sanz Moreno, J., Polo-Luque, M.-L., Garcés-Jiménez, A., 2020. Predicting Infectious Diseases by Using Machine Learning Classifiers, in: Rojas, I., Valenzuela, O., Rojas, F., Herrera, L.J., Ortuño, F. (Eds.), Bioinformatics and Biomedical Engineering, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 590–599. https://doi.org/10.1007/978-3-030-45385-5_53

[9] Han, J.J., Song, H.A., Pierson, S.L., Shen-Gunther, J., Xia, Q., 2023. Emerging Infectious Diseases Are Virulent Viruses—Are We Prepared? An Overview. Microorganisms 11, 2618. https://doi.org/10.3390/microorganisms11112618

[10] Jahangir, I., Hannan, A., Javed, S., 2018. 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. pp. 1–4.

[11] Jose, P., Rajan, N., Kommu, P.P.K., Krishnan, L., 2022. Dengue and scrub typhus co-infection in children: Experience of a teaching hospital in an endemic area. Indian J. Public Health 66, 292–294. https://doi.org/10.4103/ijph.ijph_2052_21

[12] Linsuwanon, P., Auysawasdi, N., Chao, C.-C., Rodkvamtook, W., Shrestha, B., Bajracharya, S., Shrestha, J., Wongwairot, S., Limsuwan, C., Lindroth, E., Mann, A., Davidson, S., Wanja, E., Shrestha, S.K., 2024. Estimating the Seroprevalence of Scrub Typhus in Nepal. Pathogens 13, 736. https://doi.org/10.3390/pathogens13090736

[13] Mishra, S.R., Ghimire, K., Khanal, V., Aryal, D., Shrestha, B., Khanal, P., Yadav, S., Sharma, V., Khatri, R., Schwarz, D., Adhikari, B., 2025. Transforming health in Nepal: a historical and contemporary review on disease burden, health system challenges, and innovations. Health Res. Policy Syst. 23, 61. https://doi.org/10.1186/s12961-025-01321-z

[14] Rai, V.K., Chakraborty, Santonab, Chakraborty, Shankar, 2023. Association rule mining for prediction of COVID-19. Decis. Mak. Appl. Manag. Eng. 6, 365–378. https://doi.org/10.31181/dmame0317102022r

[15] Shakil, K.A., Anis, S., Alam, M., 2015. Dengue disease prediction using weka data mining tool. https://doi.org/10.48550/ARXIV.1502.05167

[16] Shmueli, G., Bruce, P.C., Yahav, I., Patel, N.R., Jr, K.C.L., 2018. Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. John Wiley & Sons.

[17] Singh, A., Singh, D., Upreti, K., Sharma, V., Rathore, B.S., Raikwal, J., 2022. Investigating New Patterns in Symptoms of COVID-19 Patients by Association Rule Mining (ARM). J. Mob. Multimed. 19, 1–28. https://doi.org/10.13052/jmm1550-4646.1911

[18] WHO, 2024. Monitoring health for the SDGs. World Health Organization..