

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Exploring the Interrelationship of Risk Factors for Supporting eHealth Knowledge-Based System

Geletaw Sahle Tegenaw

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.75033>

Abstract

In developing countries like Africa, the physician-to-population ratio is below the World Health Organization (WHO) minimum recommendation. Because of the limited resource setting, the healthcare services did not get the equity of access to the use of health services, the sustainable health financing, and the quality of healthcare service provision. Efficient and effective teaching, alerting, and recommendation system are required to support the activities of the healthcare service. To alleviate those issues, creating a competitive eHealth knowledge-based system (KBS) will bring unlimited benefit. In this study, Apriori techniques are applied to malaria dataset to explore the degree of the association of risk factors. And then, integrate the output of data mining (i.e., the interrelationship of risk factors) with knowledge-based reasoning. Nearest neighbor retrieval algorithms (for retrieval) and voting method (to reuse tasks) are used to design and deliver personalized knowledge-based system.

Keywords: knowledge-based system, eHealth, pattern discovery, data mining, association rule

1. Introduction

In Africa, on average there are nine hospital beds per 10,000 people in comparison to the world average of 27. In sub-Saharan Africa, the physician-to-population ratio is the lowest in the world [1, 2]. Countries like Ethiopia set a strategic plan to improve access and equity to preventive, essential health interventions at the village and household levels to ensure healthcare coverage in rural areas [3, 4].

On the one hand, because of the limited resource setting, the healthcare services have not got the equity of access to the use of health services, the sustainable health financing, and the

quality of healthcare service provision. The physician-to-population ratio is below the World Health Organization (WHO) minimum recommendations [1, 2]. Still, pneumonia, diarrhea, acute upper respiratory tract infection, acute febrile illness, and malaria account for 64% under five morbidity [5].

About 68% of the country's total population living in areas at risk of malaria, 75% of the country is vulnerable to malaria (defined as areas <2000 m, those areas are fertile and suitable for agriculture and accounts for up to 17% of outpatient consultations, 15% of admissions and 29% of inpatient deaths. [6, 7]

On the other hand, it has been more than four decades the computer program reasons (e.g., MYCIN was developed in the early 1970s) and uses knowledge to assist the domain experts and minimize routine activities. To prevent and control the crisis of malaria, different scholars and responsible bodies have made remarkable efforts by conducting researches, implementing strategies, and policies [6, 7]. A predictive data mining model has been constructed using Ethiopian WHO malaria database, metrological database, and national mapping database [8]. The model is accurate to determine the occurrence of death, and it is good enough to identify the cases.

However, the system is looking a mechanism to assist routine healthcare activities, administrative and medical cost, demographic challenges, and equitable health distribution. For instance, the health extension workers (HEWs) assist peripheral health services by bridging the gap between the communities and health facilities [3, 4]. Each kebele has two HEWs responsible for providing outreach services. A kebele is the smallest governmental administrative unit and on average has a population of 5000 people. The HEWs teach the community house to house for each and every person in the kebele in order to create and promote healthy lifestyles. In all, the healthcare system is searching a mechanism for teaching, alerting, and recommendation system to support the daily routine activities.

To alleviate those issues, creating a competitive eHealth knowledge-based system (KBS) is the main goal of this work which will bring unlimited benefit in low resource setting. As a case study, we choose malaria (malaria dataset) because malaria prevention and control at the community level face numerous challenges because of the climate condition (temperature, rainfall), epidemiological, and genetic, poverty, malaria outbreak, over prescription for positive result and so on. Knowing the pattern and interrelationship of risk factors is important for supporting knowledge-based system as well as prediction of malaria death occurrences/cases. An attempt is made for exploring the degree of association between malaria risk factors (related to the malaria death occurrence and case identification). Investigating the degree of interrelationship among risk factors will have a great contribution toward eradicating the outbreak of malaria. The outcome of the study helps to mitigate the severity through investigating the association of risk factors and building of a competitive knowledge-based system.

2. Literature review

Knowledge-based systems is aimed to understand or bring human-level intelligence through simulating or acting one or more of intelligent behaviors (such as thinking, problem-solving,

learning, understanding, emotions, consciousness, intuition and creativity, language capacity, etc.). On the one hand, KBS is advantageous when there is shortage of expert, decision-making for problem-solving needs an intelligent assistant, expertise is needed to be stored for future use, and so on. On the other hand, KBS faces a lot of challenges due to the abstract nature of knowledge, limitation of cognitive science, and other scientific methods [9, 10].

Knowledge representation and inference engine are the two building blocks of KBS. The knowledge acquired from experts, documents, books, and other resources has been organized using knowledge representation. The inference engine gets the knowledge and instructs how to use the knowledge to solve problems using rule- or case-based reasoning. Rule-based reasoning is a technique that reasons out about a problem based on the knowledge that is represented in the form of rules [11]. Case-based system represents situations or domain knowledge in the form of cases, and it uses case-based reasoning technique to solve new problems or to handle new situations [12].

Knowledge-based system (in health and medical domain) has made a remarkable effect through providing a reliable diagnostic and cost-effective service. Several systems have been implemented in different medical areas like cancer therapy, infections, blood diseases, general internal medicine, glaucoma, and pulmonary function tests [13, 14]. Such systems can be designed to exhaustively consider all possible diseases in a domain, which could outperform human experts to achieve a rapid and accurate diagnosis. Integrating and updating domain knowledge with knowledge discovery are relevant to increase the interestingness and user belief (such as matching discovered pattern with existing knowledge) [15]. Integrated eHealth knowledge-based system based on acquiring health knowledge will support users in exchange of knowledge and accessibility for the users through data collection, care documentation, and knowledge extraction [16].

Following and implementing a hybrid (integrated) intelligent system for medical data classification is good to produce effective knowledge-based system [17]. A promising result is scored through integration to improve the quality of knowledge-based system [17, 18]. For instance, integrating the result (rule) of the PART classification algorithm with the knowledge-based system has delivered a favorable result for the diagnosis and treatment of visceral leishmaniasis [19]. The paper by Seera and Lim has experimented and used fuzzy min-max neural networks to learn incrementally from sample data, classification, and regression tree for prediction and random forest model to achieve high classification performance [17]. Kerdprasop and Kerdprasop also tried to automate data mining model by focusing on post data mining process to step of automatic knowledge deployment using induced knowledge and formalization classification rules [20].

However, the need to work more in providing explanatory rules and handling missing data in real-world application is expected. A mechanism for handling irrelevant rule (result) is required in case of inductive experiment system and so on. To alleviate those issues, in our case study, we tried to explore the pattern and the interrelationship of risk factors for supporting the knowledge-based system as well as prediction of malaria death occurrences/cases. The result will increase the interestingness and belief of eHealth knowledge-based system which will bring unlimited benefit in low resource setting.

2.1. Research aim and objectives

Investigating the potential applicability of exploring the interrelationship of risk factors using data mining to create a competitive eHealth knowledge-based system is the main goal of this work.

2.2. Methodology

Cross industry standard process for data mining (CRISP-DM) methodology is adopted to investigate the interrelationship of malaria risk factors. Then, to design eHealth knowledge-based system, nearest neighbor retrieval algorithms (for retrieval) and voting method (to reuse tasks) are used. The technique is easy in exploring relevant cases and provides an opportunity to retrieve partially matching cases [21–24].

The malaria data is collected from a zonal health facility in each of the 86 zones of Ethiopia. To understand the problem domain, we used observation, interviewing with experts and data managers and reviewing documents, reports, and literatures. This helps us to select and integrate decisive attributes from different sources. The data selected from the WHO (World Health Organization) database is integrated with the decisive attributes (like temperature, rainfall, and altitude) extracted from the Ethiopian National Meteorological Agency and Mapping Agency in order to find the association of risk factors.

Exploratory data analysis is performed to get familiar with the data and prepared for investigating the degree of interrelationship. The data mining task is finding the internal association between data elements that will determine the occurrence of death/case. To maintain the quality of the data, preprocessing tasks such as data cleaning (handling missing values, noisy, and outer values), data integration tasks, and data transformation tasks are performed.

The collected Ethiopian WHO malaria database contains five basic attributes (more than 37,000 records) that provide information about the geographic location and period of coverage. These attributes include country name, region (administrative regions from which malaria information is collected), zone and health facility name, year, and month. The detailed information of the attributes is categorized based on the WHO standards and explicitly represents the detailed information about malaria in each zone of the region across Ethiopia. These categories contain age (less than, equal, or greater than 5 year), malaria type (*P. vivax* and *P. falciparum*), cases (inpatient and outpatient), inpatient cases (cases and deaths), severe anemia (inpatient malaria cases less than 5 years and greater than 5 years), and uncomplicated lab-confirmed malaria less than 5 years and greater than 5 years (*P. vivax* outpatient cases and *P. falciparum* outpatient cases). Each attribute is preprocessed and statistically summarized into address, patient profile, weather, and altitude. For example, **Table 1** presented the statistical summary of uncomplicated malaria less than 5 years of lab-confirmed *Plasmodium falciparum*.

In order to extract hidden patterns and relationships within the data, from the initial dataset, a number of attributes are constructed. As shown in **Table 2**, from malaria with severe anemia, attributes such as age, malaria type, cases, and malaria visits, as well as the number of cases and deaths, are constructed.

Summary of the datasets compiled for association rule discovery with their possible nominal values and description is depicted in **Figure 1**. The malaria dataset used for this study consists

Uncomplicated Malaria <5 years P.FALCIPARUM Lab Confirmed				
Cases	Yes	No	Missing Value	Noisy Value
Frequency	105	4555	43	1
Percent	2.232%	96.832%	0.914%	0.021%

Table 1. Summary of uncomplicated malaria less than 5 years.

Initial Dataset Attribute	Target Dataset (Attribute Constructed)
Inpatient malaria with severe anemia < 5 years cases	- Age
Inpatient malaria with severe anemia < 5 years deaths	- Type of cases
Inpatient malaria with severe anemia > 5 years cases	- Type of malaria
	- Type of malaria visits
	- Number of cases
	- Number of deaths

Table 2. Malaria with severe anemia and list of attributes constructed.

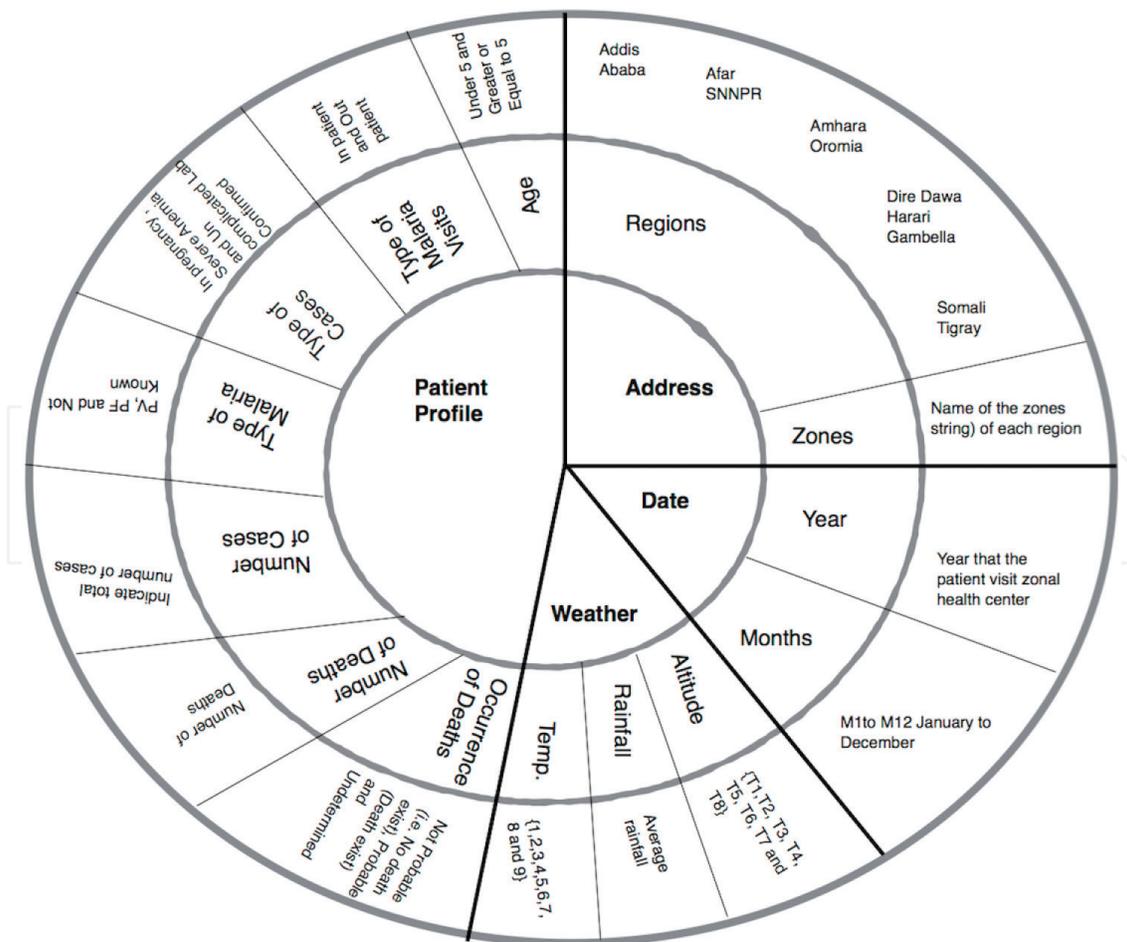


Figure 1. Summary of attributes and their transformed values.

of 14 attributes. The first inner part mainly presents the category of the attributes (i.e., profile, weather, address, and date), the second inner part presents the list of attributes constructed from category, and the last inner part indicates the value of each attribute. The region and zones contain all administrative regions in Ethiopia and the locations where the patients live. The date (year and month) indicates the year and the months that the patients visit the zonal health center. The age indicates the category of patients usually classified under 5 and greater or equal to 5. The type of malaria visits is classified into inpatient and outpatient malaria visits. The attribute type of case indicates the category of cases called in pregnancy, severe anemia, and uncomplicated lab-confirmed malaria. The type of malaria attributes contains PV, PF, and not-known values. In case of severe anemia and in pregnancy, the type of malaria is not known or determined in the dataset. A number of cases and deaths indicate the total number of cases and deaths, respectively. The attribute occurrence of deaths contains not probable (i.e., no deaths exist), probable (deaths exist), and undetermined values. The rainfall attribute contains a numeric value to represent average rainfall. In the case of outpatient visit, in pregnancy and uncomplicated lab-confirmed malaria, death is not known or listed in the dataset. The transformed temperature values 1, 2, 3, 4, 5, 6, 7, 8, and 9 represents $0-5^{\circ}\text{C}$, $5-10^{\circ}\text{C}$, $10-15^{\circ}\text{C}$, $16-20^{\circ}\text{C}$, $21-25^{\circ}\text{C}$, $25-30^{\circ}\text{C}$, $31-35^{\circ}\text{C}$, $35-40^{\circ}\text{C}$, and $>40^{\circ}\text{C}$, respectively. The transformed altitude values T1, T2, T3, T4, T5, T6, T7, and T8 represent $>3500\text{ m}$, $2500-3500\text{ m}$, $2000-2500\text{ m}$, $1500-2000\text{ m}$, $1000-1500\text{ m}$, $500-1000\text{ m}$, $0-500\text{ m}$, and $<\text{zero}$ values, respectively.

3. Apriori method

The Apriori algorithm (a well-known association rule discovery method) takes a dataset with a list of items that can be easily transformed into a transaction form by creating an item for each attribute value pair that exists in the dataset [25–27]. Minimum support and minimum confidence thresholds are also defined to enable Apriori algorithms identify frequent items that are strongly associated. **Table 3** presented the step-by-step procedure to mine and extract frequent items using Apriori methods.

Given a support threshold (S), sets of X items that appear in greater than or equal to S baskets are called frequent item sets. Find all rules on item sets of the form $X \rightarrow Y$ with minimum support and confidence. For example, if-then rules about the content of the baskets $\{i_1, i_2, \dots, i_k\} \rightarrow j$ means “if a basket contains all of i_1, \dots, i_k then it is likely to contain j .” A typical question of Apriori is to “find all association rules with support $\geq S$ and confidence $\geq C$.” In

- | |
|--|
| <ul style="list-style-type: none"> – Initially, scan database once to get frequent 1-itemset – Generate candidate $k+1$ itemsets from frequent k itemset. For each k, we construct two sets of k-tuples: <ul style="list-style-type: none"> • C_k = the set of candidate k-tuples • L_k = the set of truly frequent k-tuples. – Test the candidates against DB – Terminate when no frequent or candidate set can be generated |
|--|

Table 3. Apriori methods.

general, support of an association rule is the frequency of occurrence of the set of items it mentions, and confidence of this association rule is the probability of j given i_1, \dots, i_k . It is the number of transactions with i_1, \dots, i_k containing item j . This will measure the strength of associations between i_1, i_2, \dots, i_k , and j .

The key concepts are frequent item sets (the set of items which have minimum support, denoted by L_i for i th-item set), a priori property (any subset of frequent item set must be frequent), and join operation (to find L_k , a set of candidate k -item sets are generated by joining L_{k-1} with itself). Once frequent item sets are obtained, it is straightforward to generate association rules with confidence larger than or equal to a user-specified minimum support and minimum confidence. The next top quality of the Apriori algorithm is to implement its achievement of good performance by reducing the size of candidate sets that are considered and selected for frequent k -item set [28].

A class implementing an Apriori-type algorithm iteratively reduces the minimum support until it finds the required number of rules with the given minimum confidence [29].

For mining Weka (Waikato Environment for Knowledge Analysis), knowledge discovery tool using Java is used. In Weka 3.7.3, if class association rule (car) property is enabled, the class association is mined instead of (general) association rules. Class classification generates rules that are frequently happening to the probable occurrence of malaria cases. In many studies, associative classification has been found to be more accurate than some traditional classification methods, such as C4.5 [25]. Associative classification can search strong associations between frequent patterns (conjunctions of attribute-value pairs) and class labels. Because association rules explore highly confident associations among multiple attributes, this approach may overcome some constraints introduced by decision tree induction, which considers only one attribute at a time. In all, in association of rule mining, finding all the rules that satisfy both a minimum support and a minimum confidence threshold is important so as to generate strong and interesting rules from the frequent patterns.

4. eHealth knowledge-based system

Knowledge-based systems are computer programs that try to solve problems in a human expert-like fashion by using knowledge about the application domain (knowledge base) and problem-solving techniques (inference method). The rule-based reasoning technique can be used with other reasoning techniques in order to make a knowledge-based system more efficient. For example, case-based and rule-based reasoning can be used together. Rule-based system is an example of knowledge-based system that uses rules for knowledge representation and rule-based reasoning for reasoning techniques. The development of knowledge-based systems in medical areas has made it possible to provide reliable and thorough diagnostic services with a minimum cost. Such systems can be designed to exhaustively consider all possible diseases in a domain, which could outperform human experts to achieve a rapid and accurate diagnosis. Several systems have been implemented in different medical areas like cancer therapy, infections, blood diseases, general internal medicine, glaucoma, and pulmonary function tests [13, 14].

Figure 2 presented the detail architecture of the proposed system. In this research we tried to integrate the output of data mining (i.e., the interrelationship of risk factors) with a knowledge-based system. Apriori algorithm using CRISP-DM methodology is adopted to create the interrelationship of risk factors and used for knowledge acquisition to develop a knowledge-based system. Nearest neighbor retrieval algorithms (for retrieval) and voting method (to reuse tasks) are used to design the eHealth knowledge-based system. The knowledge is represented using “IF a certain situation holds, THEN take a particular action,” and the knowledge acquired (interrelationship of risk factors) from Apriori algorithm are rules. An

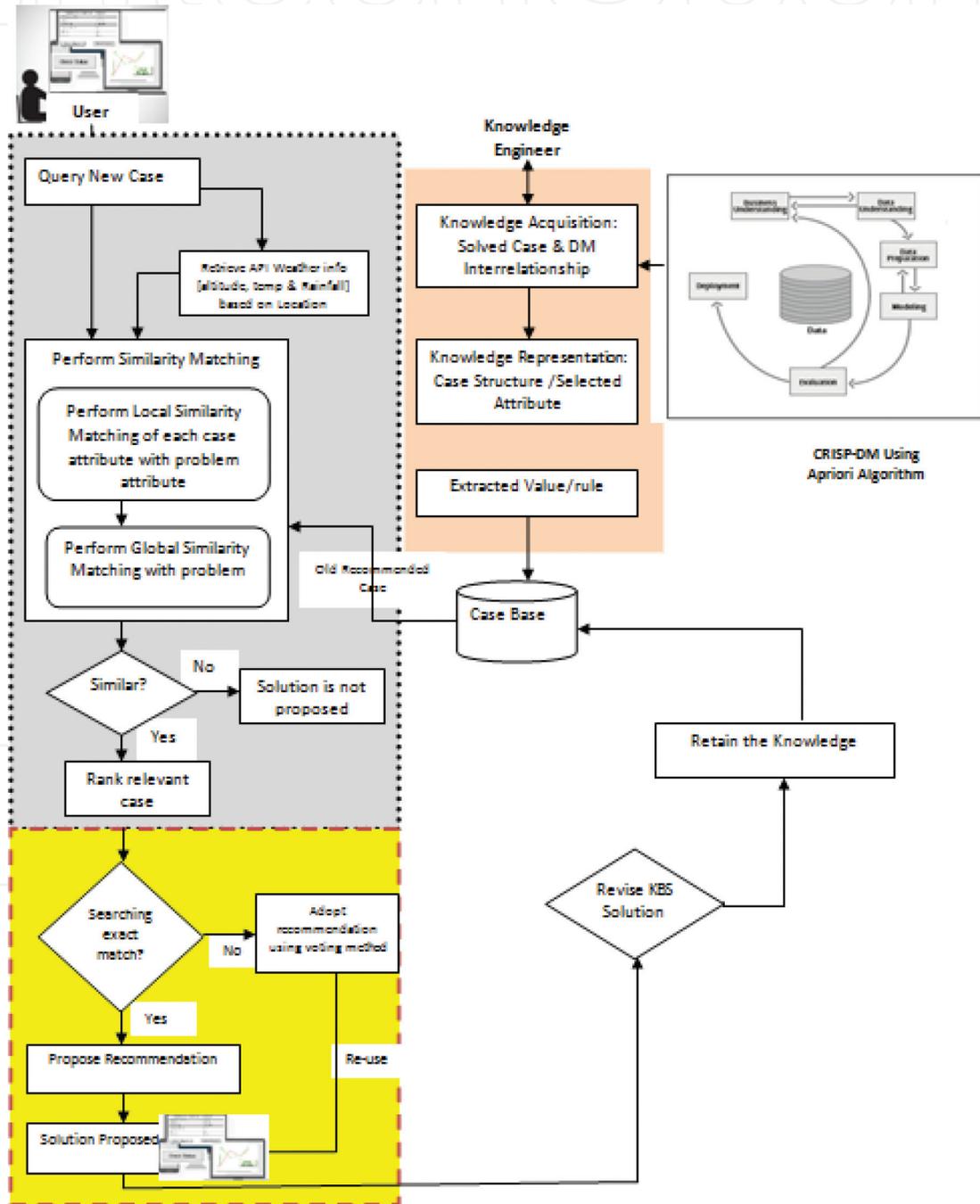


Figure 2. Adopted and proposed architecture for supporting eHealth KBS [19, 23, 24].

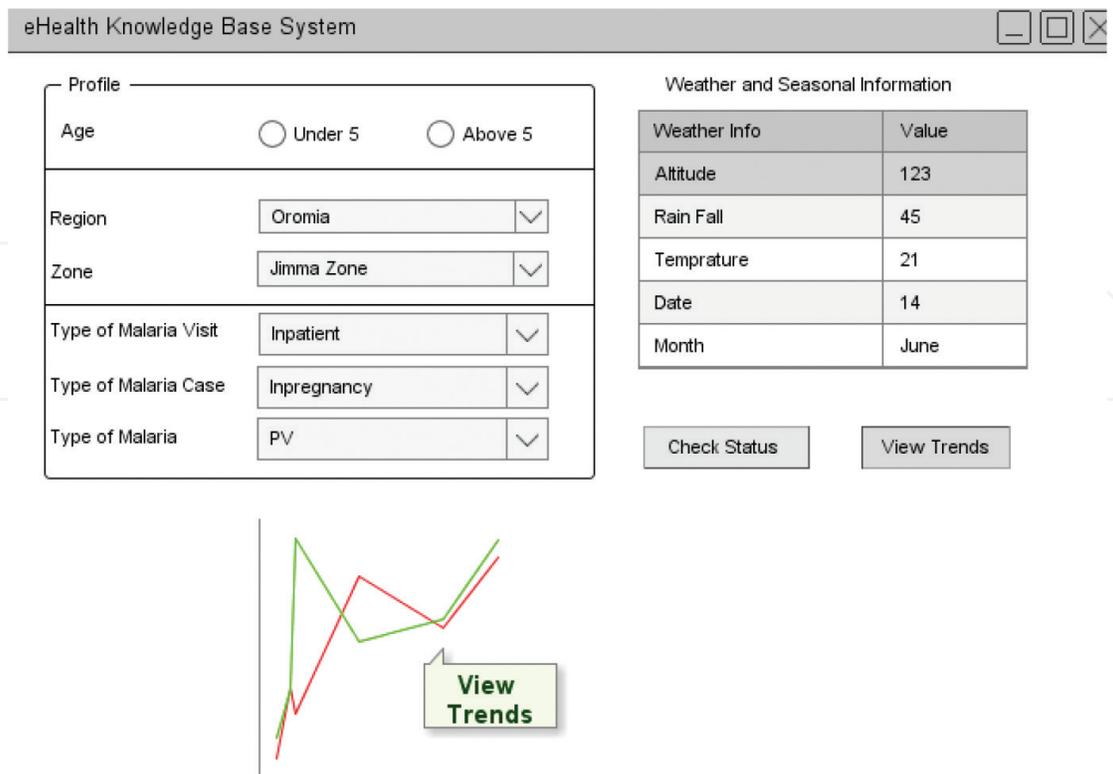


Figure 3. Graphical user interface.

inference engine of the knowledge-based system can derive the conclusion by looking the possible scenario and recommendations. The goal of the case study is to provide a personalized knowledge-based system solution using the support of interrelationship of risk factors. The user interfaces provide a communication between the user and the system.

Figure 3 presents the graphical user interface. The user tries to use the system or initiate queries by selecting profile and address information. Based on the desired location (using region and zone), weather information such as altitude, rainfall, and temperature are filled automatically from external weather API. Then, similarity matching is performed using the new queries to retrieve and recommend the proposed solution. However, if similarity matching is unsuccessful, voting technique is applied to select the relevant cases. Finally, to select or recommend the solution, the domain expert will evaluate and validate the new case. The knowledge-based system will use the validated case for future purpose.

5. Experimental results and discussions

The study tried to explore the interrelationship of risk factors for supporting eHealth knowledge-based system. We have used malaria dataset as a case study to discover the association among the various malaria risk factors using associative rule discovery data mining, and we integrate it with the eHealth knowledge-based system.

5.1. Experimental setup

A general and a class association rules are used to discover interesting association patterns. A total of 120 experiments executed using Apriori algorithm (60 experiments using general association rule and 60 experiments for class association rule mining) as depicted in **Table 4**. The confidence level is the most important parameter to attain the required objective. By considering this, the experiment is done at different confidence levels of 100, 90, 80, 70, 60, and 50%. Each confidence level is also experimented with a lower bound support of 10–100%. In both scenarios the min support of the upper bound is 100%.

Confidence level	No. of experiments	Min support lower bound	Experiment result		
			No. of association/interrelationship rules generated	No. of cycles performed	Min support used to generate the rules
100%	5	60–100%	None		
	1	50%	7	10	50%
	1	40%	7	12	40%
	3	10–30%	10	11	35%
90%	4	70–100%	None		
	1	60%	2	8	60%
	1	50%	7	10	50%
	1	40%	7	12	40%
	3	10–30%	10	13	35%
80%	4	70–100%	None		
	1	60%	2	8	60%
	5	10–50%	10	10	50%
70%	4	70–100%	None		
	1	60%	2	8	60%
	5	10–50%	10	10	50%
60%	4	70–100%	None		
	1	60%	2	8	60%
	5	10–50%	10	10	50%
50%	4	70–100%	No rule generated		
	1	60%	2	8	60%
	5	10–50%	10	10	50%
Total exp.	60				

Table 4. Scenario and result of general association rule experiment.

5.2. Generated association rules

From the experiment, we observed that the class association mining supports the rules generated in general association mining. It also discovers interesting interrelationship (with 100% confidence level) related with the type of visit, age, altitude, temperature, and malaria type. With 100% confidence level and 60% support, outpatient cases are more closely related to the undetermined occurrence of death specifically when the age group of malaria patient is greater than 5. On the one hand, the result noted that occurrence of death is mostly related to outpatient case instead of the inpatient one. This shows that health workers offer great attention and intensive care for inpatient visits. On the other hand, most of outpatient visits are uncomplicated lab-confirmed malaria. However, the occurrence of death is undetermined and probable when the type of malaria visit is outpatient and the age of the patients is greater than 5. This may be because of lack of qualified health workers and the patients are not properly prescribed as confirmed by Ndiaye et al. [30] in Senegal that the lay health workers made negative diagnostic test.

6. Discussions

Table 5 illustrates the summary of experimental results. It is difficult to determine the occurrence of death for outpatient cases, and the experimental result revealed that the occurrence of death is related with the increment of malaria case. Roca-Feltrer et al. [31] noted that the increment of malaria cases is related with the transmission intensity, seasonality, and age that lead to a probability of occurrence of deaths. For instance, the experimental result in west Gojjam (specifically in November) supports the probability of occurrence of deaths related with the increment of malaria cases.

Knowing the seasonality of malaria helps to provide proper intervention to eradicate occurrence of death and cases [31]. Roca-Feltrer et al. [32] relate the transmission intensity, seasonality, and the age pattern of malaria and confirm that younger age groups are with increasing transmission intensity. Our experimental result also confirmed that occurrence of death is undetermined if the altitude is 1500–2000 m and when the age of the person is greater than 5 with a confidence level of 62 and 60%, respectively. This happens because of high transmission intensity. With 100% confidence level, if the type of malaria visit is outpatient and age is below 5, it is difficult to predict occurrence of deaths. Interestingly, occurrence of malaria death is related with severe anemia rather than pregnancy. As discussed by Knoblauch et al.

-
- If temperature is between 15 and 200°C, the type of malaria visits is outpatient, and the type of cases is uncomplicated lab-confirmed malaria, then the occurrence of death is undetermined.
 - If the type of malaria is PV, then occurrence of death is undetermined, and if the type of malaria is PF, then occurrence of death is undetermined.
 - If age is under 5 years and the type of malaria visits is outpatient, then occurrence of death is undetermined.
-

Table 5. Summarized experimental rules.

[33], anemia is prevalent in the 6- to 59-month-old children, and the association of anemia with a child age, underlying with iron requirements, is related to growth rate, and hence iron demand declines with age. Further, the algorithm associated (with 100% confidence level) with the type of malaria is unknown for inpatient malaria visits.

However, some unexpected or interesting interrelationship is prevailed such as with 100% confident level, and it is difficult to determine the occurrence of death for both PV and PF malaria types. This needs further investigation to verify whether it is unrelated or expected.

In all, the study presents the association of malaria risk factors using climate, elevation, location, type of malaria, type of malaria visits, number of cases, and death attributes. Both general and class association minings are done using Apriori techniques for discovering the association or patterns between the occurrence of deaths with the type of cases and malaria visits. And then, integrate the output of data mining (i.e., the interrelationship of risk factors) with knowledge-based reasoning. Nearest neighbor retrieval algorithms (for retrieval) and voting method (to reuse tasks) are used to design and deliver personalized knowledge-based system.

6.1. Evaluations

The evaluation of the result is executed by combining both an expert and testing tool approaches. An overall measure of pattern values, combining novelty, usefulness, and simplicity, to achieve a predefined goal is evaluated to measure the interestingness of the interrelationship. We have used different multitudes of measurement in the evaluation such as accuracy, support level, confidence, confidence level, and complexity with a 10fold cross validation. We adopted the four measures such as sensitivity, specificity, prediction accuracy, and precision to evaluate the correctness of interrelationship and validate the system through performance testing. Tenfold cross validation is used in the experiment to predict the error rate [34, 35]. The basic measure is accuracy, which computes the percentage of correctly classified instances in the test set. The accuracy of a test compares how close a new test value is to a value predicted by if-then rules [36].

The interrelationship of risk factors (association rules) was evaluated in terms of the number of rules and meaning of patterns generated at different minimum support and confidence thresholds for measuring interestingness of the rules. Association (interrelationship) was analyzed in terms of different criteria. The criteria include the number of rules generated at different minimum support and confidence thresholds. The minimum support and confidence thresholds varied from 0.1 to 1 and 0.5 to 1, respectively. As depicted in **Table 4**, at 90% confidence level with min support of 60 and 20%, the techniques generate 2 and 10 rules, respectively. Furthermore, we investigate the following indicators of the quality of the rule ranking induced by the interestingness measures of the mining algorithm in the average rank of the first rule that covers a test instance and the average rank of the first rule that covers and correctly predicts a test instance.

The performance of eHealth knowledge-based system prototype is evaluated using test case. Thus, the effectiveness of the retrieval process of eHealth knowledge-based system reasoning is measured by using recall and precision. Precision and recall are useful measures of retrieval performance [34]. Recall is the percentage of relevant cases for the query (new case) that are

retrieved, whereas precision is the percentage of retrieved cases that are relevant to the query [34, 36, 37]. Accuracy is used to measure the performance of the reuse process [34, 36].

The case similarity testing shows that when the query is made up of attribute values that have the same value with the case from the case base, the result of the global similarity becomes 1.0. But when there is a difference in the attribute values of the query and the case in the case base, the global similarity value decreases. Therefore, adding cases in the case base improves the performance of knowledge-based reasoning system in solving problems (new cases).

The nearest neighbor algorithm, which is used to develop the retrieval process of the prototype, uses distance to compute the similarity between the query and cases by representing the cases in N dimension vector. However, the recommendation doesn't have clear boundaries as it has subjectivity and depends on the experience of the domain experts as tested and adopted in [19, 23, 24]. In addition, the importance value that is assigned to the attributes of the case structure is done manually with the help of the domain experts, as there is no research that is conducted for the importance value of the attributes in malaria case management. This could affect the result of the retrieval and the reuse performance of the prototype. However, it needs user acceptance testing (using measuring usability with the system usability scale) in real-world scenarios to measure whether the potential users would like to use the proposed system frequently or not. So that, eHealth knowledge-based system for retrieving relevant cases and proposing solution will attain promising user acceptance, accuracy, and domain expert evaluation.

7. Conclusion and future work

The experimental result presents the association of risk factors (with relation to the malaria occurrence of death and type of case identification in Ethiopia) using climate, elevation, location, type of malaria, type of malaria visits, number of cases, and death attributes. Both general and class association minings are done using Apriori techniques for discovering the association or patterns of risk factors. The results noted the existence of strong association between occurrence of deaths, type of malaria visits, age, and type of cases. More interestingly, it discovers occurrence of malaria deaths, which are mostly related with severe anemia cases rather than pregnancy. It is also important to precede usability and user acceptance testing of eHealth knowledge-based system in real time and perform testing to compare and contrast with domain experts. So, health institutions have to give great attention to provide the necessary diagnosis and treatment for anemia, especially in regions that are more vulnerable for malaria. It also provides a significant contribution to design an optimal strategy in support of malaria prevention and control program within the country.

Acknowledgements

First and foremost I would like to thank God and the Holy Mother. Then, I would like to express special thanks for Worku Birhane who provided me proofreading and valuable comments.

Author details

Geletaw Sahle Tegenaw

Address all correspondence to: gelapril1985@gmail.com

Faculty of Computing, Jimma Institute of Technology, Jimma University (JU), Ethiopia

References

- [1] ESTA, OSHD. Health in Africa Over the Next 50 years. 2013. www.afdb.org
- [2] AfricaNext Investment Research. The future of African broadband: Economics, business models and the rise of 3G. 2010;(5)
- [3] Federal Ministry of Health. Health Extension Program in Ethiopia. Addis Ababa, Ethiopia: Health Extension and Education Center; 2007
- [4] Family Health Department. National Strategy for Child Survival in Ethiopia. Addis Ababa, Ethiopia: Federal Ministry of Health; 2005
- [5] <http://www.moh.gov.et/factsheets> [Accessed: August 2017]
- [6] Federal Democratic Republic of Ethiopia Ministry of Health. Ethiopia National Malaria Indicator Survey 2007 Technical Summary; 2008
- [7] Ministry of Health. Malaria Prevention Control Program. 2014. <http://www.moh.gov.et/malaria> [Accessed: June 2014]
- [8] Sahle G, Meshesha M. Uncovering knowledge that supports malaria prevention and control intervention program in Ethiopia. *Electronic Journal of Health Informatics*. 2014;**8**(1):e7. www.eJHI.net
- [9] Sajja PS, Akerkar R. Knowledge-based systems for development. *Advanced Knowledge Based System: Model, Application & Research*. 2010;**1**:1-11
- [10] Tan C. A prototype of knowledge-based system for fault diagnosis in automatic wire bonding machine. *Turkish Journal of Engineering and Environmental Sciences*. 2008;**32**:235-244
- [11] Hayes-Roth F. Rule-based systems. *Communications of ACM*. September 1985;**28**(9):921-932. DOI: 10.1145/4284.4286
- [12] Engelmores RS, Feigenbaum E. Knowledge-Based Systems in Japan. Japanese Technology Evaluation Center. Maryland: Loyola College; 1993. <http://www.wtec.org/loyola/kb/toc.htm>
- [13] Abdelhamied K, Hafez S, Abdalla W, Hiekal H, Adel A. A rule-based expert system for rapid problem solving in crowded outpatient clinics in Egypt. *IEEE*. 1988;**3**:1419-1420

- [14] Pandey J, Bajpai D. Developmental design of a rule based expert system for diagnosis. In: Proceedings of the First Regional Conference, IEEE Engineering in Medicine and Biology Society and 14th Conference of the Biomedical Engineering Society of India. New Delhi: An International Meet; 1995. pp. 1/41-1/42. DOI: 10.1109/RCEMBS.1995.508680
- [15] Pohle C. Integrating and updating domain knowledge with knowledge discovery. In: 6th International Conference for Business Informatics 2003 (WI-2003); Dresden, Germany; September 15-17; 2003
- [16] Nasiri S, Fathi M. Toward an integrated e-health based on acquired healthcare knowledge. In: 2014 Middle East Conference on Biomedical Engineering (MECBME); 2014
- [17] Seera M, Lim CP. A hybrid intelligent system for medical data classification. *Expert Systems with Applications*. 2014;**41**:2239-2249
- [18] Sedighian Z, Javanmard M. The effect of data mining on expert systems used for improving efficiency of correct speech E-learning systems. *Advances in Natural and Applied Sciences*. 2014;**8**(10):102-106
- [19] Mulugeta T, Beshah T. Integrating Data Mining Results with the Knowledge Based System for Diagnosis and Treatment of Visceral Leishmaniasis. 2015. Available at www.ijarcsse.com
- [20] Kerdprasop K, Kerdprasop N. Bridging data mining model to the automated knowledge base of biomedical informatics. *International Journal of Bio-Science and Bio-Technology*. 2012;**4**(1):13
- [21] Martin B. Instance-based learning: Nearest neighbour with generalisation [MSc thesis]. University of Waikato; 1995. Available at: <http://www.cs.waikato.ac.nz/pubs/wp/1995/uow-cs-wp-1995-18.pdf> [Accessed: March 27, 2011]
- [22] Mishra D, Sahu B. Feature selection for cancer classification: A signal-to-noise ratio approach. *International Journal of Scientific & Engineering Research*. Hamilton, New Zealand: University of Waikato. 2011;**2**(4):1-7
- [23] Salem AM. Case-based reasoning technology for medical diagnosis. *World Academy of Science, Engineering and Technology*. 2007;**31**:9-13
- [24] Bekele H. A case based reasoning knowledge based system for hypertension management [MSc thesis]. Ethiopia: Addis Ababa University; 2011
- [25] Han J, Micheline K. *Data Mining: Concepts and Techniques*. Waltham, USA: Morgan Kaufman Publishers is an imprint of Elsevier; 2006
- [26] Larose DT. *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey, Canada: John Wiley & Sons; 2005
- [27] Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems. 2nd ed. Physical Description xxxi. San Francisco: Morgan Kaufman Publishers; 2005. 525 p. ill. <https://trove.nla.gov.au/version/46617902>

- [28] Wu X, Kumar V, Ross Quinlan J, et al. Knowledge and Information Systems. Springer-Verlag. 2008;**14**:1. <https://doi.org/10.1007/s10115-007-0114-2>. Print ISSN: 0219-1377. Online ISSN: 0219-3116
- [29] The University of Waikato, WEKA Manual (Waikato Environment for Knowledge Analysis) for Version 3-7-4, This manual is licensed under the GNU General Public License version 2. Available at <http://www.gnu.org/copyleft/gpl.html>
- [30] Ndiaye Y, Ndiaye JLA, Cisse B, Blanas D, Bassene J, Manga IA, Ndiath M, Faye SL, Bocoum M, Ndiaye M, Thior PM, Sene D, Milligan P, Gaye O, Schellenberg D. Community case management in malaria: Review and perspectives after four years of operational experience in Saraya district, south-East Senegal. *Malaria Journal*. 2013;**12**:240
- [31] Roca-Feltrer A, Armstrong Schellenberg J, Smith L, Carneiro I. A simple method for defining malaria seasonality. *Malaria Journal*. 2009;**8**:276
- [32] Roca-Feltrer A, Carneiro I, Smith L, Schellenberg JRMA, Greenwood B, Schellenberg D. The age patterns of severe malaria syndromes in sub-Saharan Africa across a range of transmission intensities and seasonality settings. *Malaria Journal*. 2010;**9**:282
- [33] Knoblauch AM, Winkler MS, Archer C, Divall MJ, Owuor M, Yapo RM, Yao PA, Utzinger J. The epidemiology of malaria and anaemia in the Bonikro mining area, central Côte d'Ivoire. *Malaria Journal*. 2014;**13**:194
- [34] McSherry D. Precision and recall in interactive case-based reasoning. In case-based reasoning research and development (ICCBR). *Lecture Notes in Artificial Intelligence*. 2001;**2080**:392-406
- [35] Kohavi R. A study of cross validation and bootstrap for accuracy estimation and model selection. In: *The International Joint Conference on Artificial Intelligence*; 1995
- [36] Junker M, Hoch R, Dengel A. On the evaluation of document analysis components by recall, precision, and accuracy. In: *Proceedings of the Fifth International Conference on Document Analysis and Recognition*; 1999. pp. 713-716
- [37] Losee RM. When information retrieval measures agree about the relative quality of document rankings. *Journal of the American Society for Information Science*. 2000;**51**: 834-840