Article

Not peer-reviewed version

E_Apriori: An Efficient Machine Learning Algorithm for the Control of Malaria

Gbenga Adegbite *

Posted Date: 21 August 2023

doi: 10.20944/preprints202308.1388.v1

Keywords: Malaria; Machine Learning; Demographic and Health Survey; biological insights; control interventions

Article

E_Apriori: An Efficient Machine Learning Algorithm for the Control of Malaria

Gbenga Adegbite [1,2]

[1] Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria; gbenga.adegbite@stu.cu.edu.ng

[2] Covenant University Bioinformatics Research, College of Science and Technology, Covenant University, Ota, Nigeria

Abstract: The discovery of interesting inter-relationships between the different malaria epidemiological parameters is essential towards the disease control. However, existing associative rule-based machine learning algorithms for pattern discovery are slow while working on high-dimensional Malaria Indicator Survey (MIS) data, with the further challenge of data under fitting and inadequate result visualization. Hence, this work proposed a novel and efficient associative rule-based machine-learning algorithm with enhanced graphical visualization capacity for rigorous and confident biological result interpretation for malaria control. Through empirical and asymptotic comparative time-complexity performance evaluations, the proposed algorithm scaled better than other existing associative rule-based machine learning algorithms while maintaining its accuracy. The algorithm was applied to two real MIS data sets obtained from the Demographic and Health Survey repository and other supplementary literature source using Nigeria as a case study. The resulting interesting malaria epidemiological discovered novel trends were: a) the malaria disease might not be associated with the anemia symptom; b) there was no significant association between the anemia symptom and the wealth indices of individuals; c) there were other parameters associated with the insecticide resistance capacity of the malaria vector asides the knock down resistance alleles; d) the population dynamics of the malaria vector was not associated with the malaria disease endemicity. In conclusion, this work developed a computationally efficient and user-friendly associative rule-based machine-learning algorithm called E_Apriori for the control of the malaria disease.

Keywords: malaria; machine learning; Demographic and Health Survey; novel trends; control

Introduction

Malaria is often endemic in tropical Sub Saharan Africa [1–4]. However, children and pregnant women are more susceptible to malaria than other class of individuals [5–9]. This is due to an underdeveloped and suppressed immune system for children (0–4 years), and pregnant women, respectively [10,11]. Hence, malaria epidemiologists and control program formulators preferentially prioritize these groups of individuals (children and pregnant women) during Malaria Indicator Survey (MIS), and its implementation, respectively [10,12–15].

MIS is a field-data collection study, which is often targeted towards the malaria endemicity peak periods in a country [10,12–14,16]. However, the malaria endemicity peak periods are often associated with rainfall and temperature climatic factors [8]. Moderately high rainfall boosts the prevalence of malarial mosquito breeding reservoirs and sites, leading to the imminent malaria endemicity during this period [17–19]. Often, moderately high temperature increases the malaria transmission rate by enhancing the *Plasmodium* parasite development into the infective sporozoite stage in the malarial vector, at a faster rate [19,20]. Hence, experimental biologists have established and supported the hypothetical inference that the high prevalence of mosquitoes in a region, contributes to the endemicity of malaria in that region, which this study sought to investigate.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2023 doi:10.20944/preprints202308.1388.v1

Insecticide resistance as an important MIS data field, provides information about the rate at which the malarial vector develop resistance to insecticides such as pyrethroids which can be used to treat mosquito net towards malaria control [21]. Further, the knock down resistant (kdr) allele describes the presence of an abnormal genetic make-up in the malarial vector caused by mutations in the nervous system of the vector, which has being biologically validated to enhance the insecticide resistance capability of the vector [22,23]. However, this study further sought to investigate these two variable relationships.

Anemia is the major symptom associated with malaria in Tropical Sub-Saharan Africa when a mosquito infected with any of the *Plasmodium* species (*P. falciparum*, *P. ovale*, *P. vivax* and *P. malariae*) bites a susceptible human, to reduce the level of Red Blood Cells [24]. Hence there have being an existing hypothetical inference that most individuals with the anemia symptom are expected to successfully accept malaria treatment [24] which this study investigated. Further, the prevalence of the anemia symptom among low-income individuals [21,25] was a research question that this study sought to answer.

Sporozoite rate is useful to determine the rate of malarial *Plasmodium* parasite infection in the female *Anopheles* vector host blood cells [26,27]. It is expected that the number of susceptible mosquitoes in a region will determine the sporozoite rate in any given geo-referenced location and vice-versa [26,27]; which might not be a consistent hypothetical inference.

Hence, there is the need to develop computationally efficient algorithms to discover and deduce novel malaria epidemiological trends from clinical MIS data towards proposing malaria control interventions. In furtherance of this research attempt, this study performed a critical computational appraisal of existing machine learning algorithms. However, the Apriori rule-mining machine-learning algorithm [28] was of primary interest because of its existing known utility in associative rule mining and interesting pattern discovery while working on biological and clinical data.

However, the major challenge with the Apriori algorithm [28] is that it is computationally slow while working on high-dimensional biological data towards generating associative patterns and may truncate the process of pattern discovery before completion [29]. In addition, there is the challenge of data under fitting and inadequate biological result visualization [29,30]. Hence, this work proposed an improved computationally efficient but cost-effective machine-learning algorithm called E_Apriori that will be useful in generating accurate but simple to interpret associative patterns, from high-dimensional biological and clinical data towards malaria control.

Methods

This work was based on extending the existing Apriori association rule-based machine-learning algorithm towards achieving better computational efficiency. We first describe the conventional Apriori association rule-based machine-learning algorithm by presenting its procedure in the next section. An extended Apriori algorithm (E_Apriori) was then presented in the subsequent section.

Basic Model

The basic model for this study was the Apriori associative rule-based machine-learning algorithm [28]. The problem that the Apriori algorithm sought to solve was stated mathematically using biological conceptualizations as follows. Given that B is a set of MIS biological data representations ($B_{k\infty}$) such that k is the number of subset representations contained in every unique parent data field. Hence, the association rule is generated for B_1 and B_2 ($B_1 \rightarrow B_2$) when it is assigned a unique Biological combination ID (BID) given that $B_1 \cap B_2 \neq 0$ with a user defined minimum support for a confident biological result generation.

Hence, the detailed pseudocode for the conventional Apriori Algorithm [28] was given as follows.

1. Input B = $\{B_{k\infty}\}$
2. For (k=2; $B_{k-1} \neq 0$; k++) do begin
3. C_k = apriori-gen (B_{k-1}) // C_k : biological candidates

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2023 doi:10.20944/preprints202308.1388.v1

4. For all biological combinations C_{BID} = subset (C_k, BID)
5. For all biological candidate B_c € C_{BID}
6. B_c.count++
7. B_k = {B_c £ C_k ׀ B_c.count ≥ min_support}
8. Output U_k B_k

The E_Apriori Machine Learning Algorithm

From the biological emphatic point of view, the novel E_Apriori was proposed to work on high-dimensional MIS data towards overcoming the challenge of data under fitting, visualization and slowness. Hence, the procedure of the proposed novel E_Apriori algorithm was formally described to prove its correctness as follows.

Given a high-dimensional Biological MIS data input (B) of size D, with $N_\infty \leq 10000$ given that N is the number of data fields and $B_{k\infty}$ is used to represent the number of possible subsets in every given data field such that $k_\infty \leq 9$ for k > 0 & k = 10; then find the frequent associative value representations (f) of a unique candidate biological data field using the set union representation ($B_{k\infty}$ {k_1 U $k_{2...k9}$}) with a user-defined minimum support (recommended min_support ≥ 70) stored in an array output file (L_F) of Index: 1_1. Repeat for the next unique biological candidate data input of $M_{k\infty}$ to find the frequent associative value representations (f), by using the set union $M_{k\infty}$({ k_1 U $k_{2...k9}$}) with a user-defined minimum support (recommended min_support ≥ 70) stored in an array output file (M_F) of Index: 1_2. Then find the frequent associative value representations between a unique biological candidate input (L_F) and another unique biological candidate input (M_F) by using the set union given as M_F U L_F with a user-defined minimum support(recommended min_support ≥ 70) to be stored in a matrix data structure of Index I_RI_C. The association between M_F and L_F mathematically as B_k (L_F →M_F) such that min_support (((L_F U M_F)/L_F) which is stored in a C_t data file as the output. Further, a graphical visualization is provided for the resulting associative output as a graph (set x-axis: L_F $ set y-axis: M_F).

The pseudocode for the Novel E_Apriori algorithm was given further as follows.

$$\text{Problem: } \arg\min_t (B_k (L_F \rightarrow M_F)) \quad (1)$$

$$\arg\max_a (B_k (L_F \rightarrow M_F)) \quad (2)$$

B_k: The total number of subset representations contained in every unique MIS data field.
L_F: A unique candidate MIS data field
M_F: Another unique candidate MIS data field
min_t: minimize computational time
max_a: maximize computational accuracy
arg: argument
→: association

Input: B_∞ {D, N, k}
B_∞: MIS data
D: size of the MIS data
N: number of unique data field representations in the MIS
k: number of subset representations in any given candidate data field.

Output
B_k(L_F→M_F)
Plot_graph (set x-axis: L_F $ set y-axis: M_F)

1) function E_Apriori (B_∞ {D, N, k})
2) Given that data field ($B_{k\infty}$ € N such that $N_\infty \leq 10000$)
3) For (k=10, $B_{k-1} \neq \emptyset$, k++) {
4) Prune (B_{K-1})

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2023 doi:10.20944/preprints202308.1388.v1

5) Find frequent (f) biologically correlated values for a candidate biological data field (L)
6) $L_F = (B_{k\infty}\{k_1 \cup k_{2\ldots k9}\}) \geq min\_sup$
//Repeat step 3,4,5,6 for the next data field M
7) Given that $M_{k\infty} \in N$ such that $N_\infty \leq 10000$
8) For ($k=10$, $M_{k-1} \neq \emptyset$, k++) {
9) Prune (M_{K-1})
10) Find frequent (f) biologically correlated values for a candidate biological data field (M)
11) $M_F = (M_{k\infty}\{K_1 \cup K_{2\ldots K9}\}) \geq min\_sup$
12) $I_R I_C$ = find frequent association ($L_F \cup M_F$)
13) $B_k (L_F \rightarrow M_F) = min\_support ((L_F \cup M_F)/L_F)\}$ //Output of $B_k (L_F \rightarrow M_F)$ saved in C_t.
14) Output Plot_graph (set x-axis: L_F $ set y-axis: M_F)
15) Output $I_R I_C$, B_k
16) Repeat all steps for all new B_k candidates contained in B_∞
17) End

Figure 1 shows the basic flow chart of the proposed E_Apriori Algorithm.

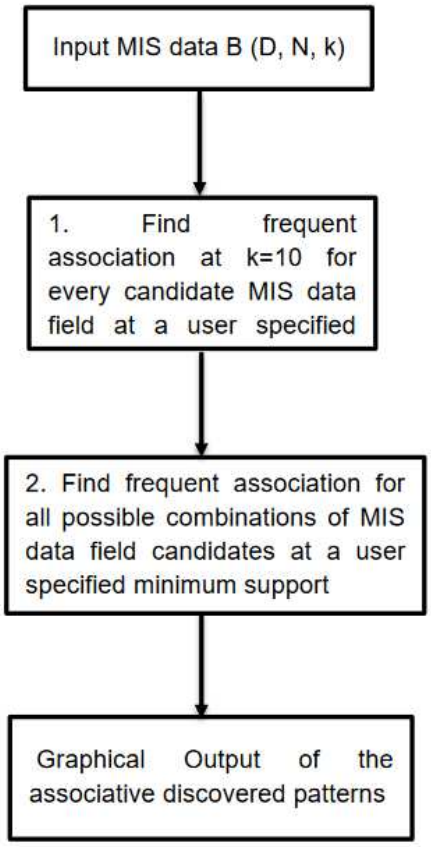


Figure 1. The Basic Flow Chart of the Proposed E_Apriori Algorithm.

Comparative Performance Evaluation of the E_Apriori Algorithm

The comparative asymptotic time complexity analysis of the novel E_Apriori algorithm was done against the existing Apriori algorithms [28,31–37] to assess its computational efficiency. The average empirical time-complexity comparative performance evaluation of the proposed E_Apriori algorithm against the existing Apriori algorithms [28,38] was done on a conventional sequence-based computer of Core i5 Processor, 8GB RAM and Windows 10 Operating System specification using the high-dimensional Nigerian-georeferenced MIS 2015 data.

To further confirm the associative pattern discovery accuracy of the proposed algorithm, the empirical comparative performance analysis of the E_Apriori algorithm against other existing

 doi:10.20944/preprints202308.1388.v1

Apriori algorithms was done on a parallel computing facility with ten (10) Central Processing Units (CPUs) using the same high-dimensional Nigerian georeferenced MIS 2015 data.

Implementation of the E_Apriori Algorithm towards Associative Pattern Discovery

The E_Apriori algorithm was applied extensively to the MIS data sets, which were described as follows. The MIS dataset for Nigeria used in this study, are from two major secondary sources: the Demographic and Health Survey (DHS) Repository and other supplementary academic publication [39].

This work assessed the MIS 2015 high-dimensional Nigerian geo-referenced dataset from the DHS Repository. The MIS 2015 data comprises of Seven Thousand, Seven Hundred and Forty-Five (7745) observations for every corresponding variable of Two Thousand, Six Hundred and Forty-nine (2649) representations. In addition, there is a minimum subset of nine representations for every unique variable. Further, this study assessed other Nigerian geo-referenced dataset from supplementary literature source [39]. The supplementary literature-sourced data comprises of Six Hundred and Thirty-three (633) observations with thirty-four (34) unique variables. However, this study reported only the MIS variable observations where the novel malaria epidemiological trends were discovered.

The following numerical value-labeled codes and its respective semantics were used to represent the different variant categorical observations of anemia: 1- severe anemia, 2- moderate anemia, 3- mild and 4- not anemic. Further, the following numerical value-labeled codes and its respective semantics were used to represent the different variant categorical observations of wealth index: 1- poorest, 2-poorer, 3-middle, 4- richer, and 5-richest. Also, the following numerical value-labeled codes and its respective semantics, were used to represent the different variant categorical observations of malaria treatment acceptance: 1- accepted malaria treatment, 2- refused malaria treatment, and 6- other treatment acceptance.

Binary logistic variable observations of 0 and 1 were used as program codes to represent susceptible and resistant kdr alleles, respectively; which was same for the insecticide resistance data field. Further, the sporozoite rate were represented as percentile variables while the number of mosquitoes were of discrete data observation type.

This study followed the data preprocessing de-facto pipeline of outlier detection and removal of null values [40] before the further analytics of discovering novel associative malaria epidemiological patterns using the novel E_Apriori machine-learning algorithm proposed in this study.

3. Results of the E_Apriori Algorithm Comparative Performance Evaluation

The asymptotic time-complexity analysis of the novel E_Apriori algorithm was expressed mathematically as follows when it is given that k is the number of subset representations contained in all the parent MIS data field variables.

$$f(k) = (3k^2 + 5k + 1)^{-1} \quad (1)$$

$$f(k) \geq 3k^{-2} \quad (2)$$

$$g = k^{-2} \quad (3)$$

$$f(k) \geq 3g \text{ for all values of } k > 0 \quad (4)$$

Hence, the asymptotic time complexity of the novel E_Apriori algorithm was O (g) = O (k^{-2}), which was better than other existing Apriori algorithms when discovering interesting association patterns as shown in Table 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2023 doi:10.20944/preprints202308.1388.v1

Table 1. Asymptotic Time Complexity Comparative Performance Evaluation of E_Apriori Algorithms against existing Apriori Algorithms.

Time taken by other known Apriori Algorithms	Time taken by the novel E_Apriori algorithm	References
$O(Dk^2)$	$O(k^{-2})$	[31,32]
$O(Dk)$	$O(k^{-2})$	[33,34]
$O(Dk\|C_k\| + t\|L_{k-1}\|\|L_{k-1}\|)$	$O(k^{-2})$	[28]
$O(D\text{-}k^2)$	$O(k^{-2})$	[35]
$O(k\|L_{k-1}\|\|L_{k-1}\|)$	$O(k^{-2})$	[36]
$O(k)$	$O(k^{-2})$	[37]

The average sequential empirical running time of the proposed E_Apriori, T_Apriori, and the conventional Apriori algorithms, were 60s, 360s and 560s, respectively as shown in Figure 2. Further, the E_Apriori successfully finished generating the association patterns without truncating before completion to overcome the challenge with the T_Apriori and the conventional Apriori algorithm during the sequential computing experimentation.

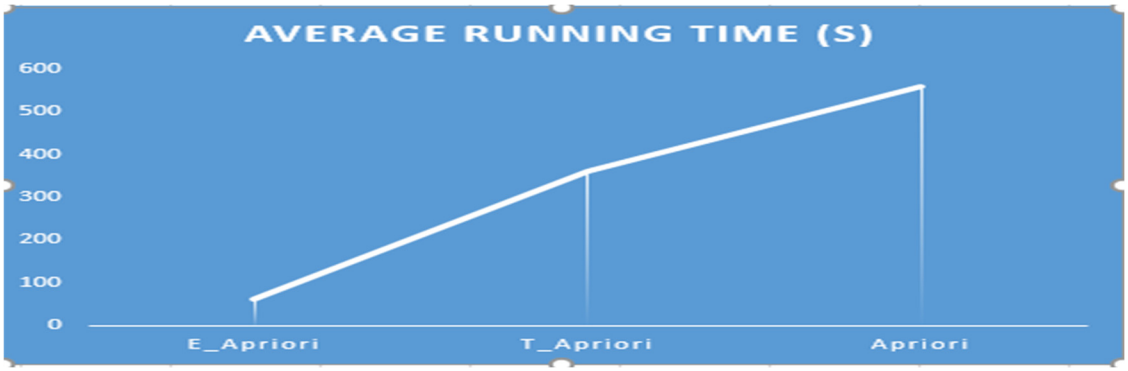


Figure 2. Average algorithmic time-complexity comparative performance evaluation.

To confirm the associative pattern discovery accuracy of the proposed algorithm, the E_Apriori algorithm generated the same associative pattern results as the conventional Apriori and T_Apriori algorithms during experimentation on parallel systems. During the parallel system experimentation, all the machine-learning algorithms (E_Apriori, T_Apriori and the conventional Apriori) successfully completed the associative pattern generation.

The Novel Malaria Epidemiological Trends and Patterns

Majority (75%) of the poorest individuals (1) were anemic which was the same for the poorer (2), middle (3), richer (4), and richest (5) individuals. Further, minority (25%) of the poorest individuals (1) were not anemic which was same for the poorer (2), middle (3), richer (4), and richest (5) individuals. Confidently, there was no significant difference in the anemia and non-anemia cases when the wealth indices of the sample individuals were at the poorest (1), poorer (2), middle (3), richer (4), and richest (5) levels as visualized using a scatterplot diagram shown in Figure 3. This emphasized that there was no evidence of randomness and linearity in the wealth indices to anemia distribution ratio.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2023 doi:10.20944/preprints202308.1388.v1

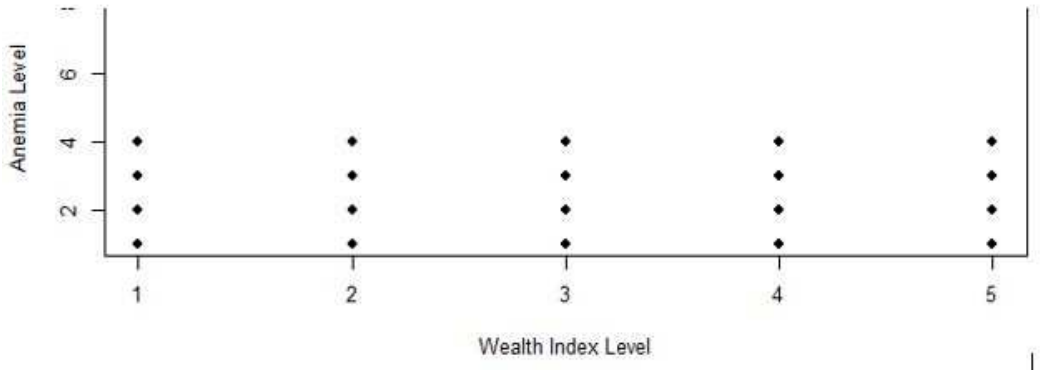


Figure 3. Association between the Anemia and Wealth Indices.

Figure 4 shows the associative relationship between the anemia and accepted malaria treatment data fields. 40% of sampled individuals with mild and moderate variants of anemia successfully accepted malaria treatment whereas 20% refused the malaria treatment and other forms of treatment, despite being anemic. Also, 20% of those that are not anemic, accepted the malaria treatment while another 20% of anemic sampled individuals accepted different forms of treatment against other diseases but refused the malaria treatment..Hence, the anemia to malaria treatment followed a random distribution.

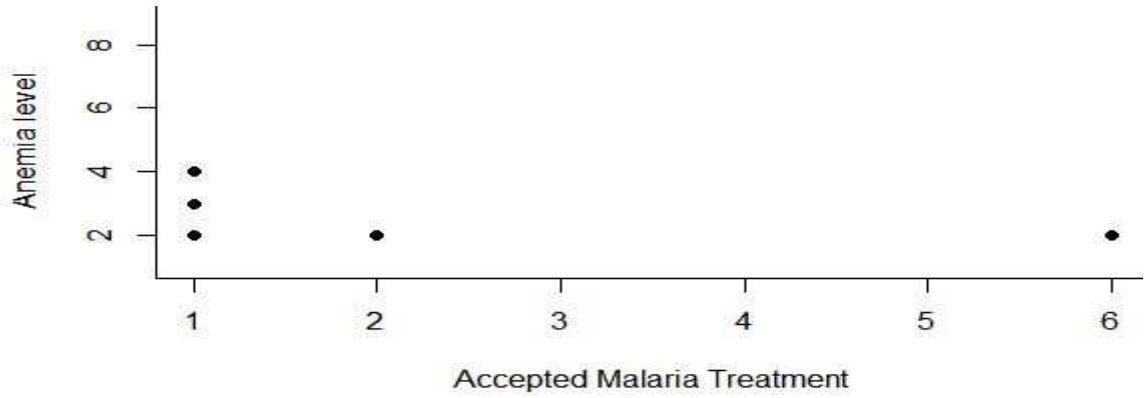
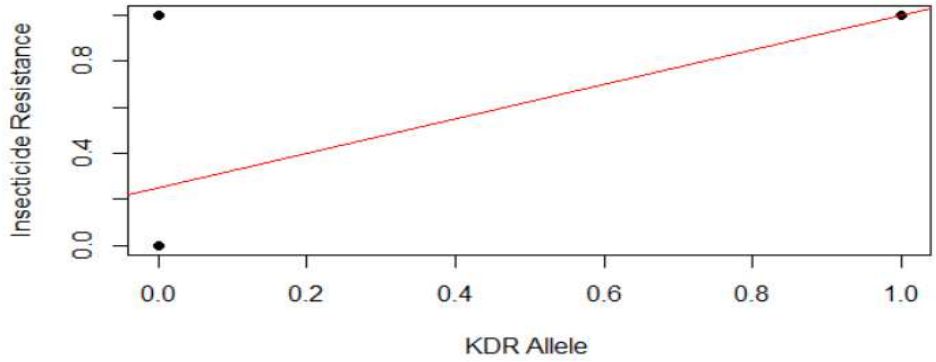


Figure 4. Association between Anemia Level and Accepted Malaria Treatment.

Further, the results of the relationship between insecticide resistance level of the malarial vector and the kdr alleles present in its genetic makeup showed that there was a weak positive association between these variables based on the linear-regression deductive method. From the results shown in Figure 5, it was also evident that 33.3333% of the malarial vector were not resistant (susceptible) and lack the kdr alleles. Further, 33.3333% of the malarial vector were resistant to insecticides while exhibiting the kdr alleles' presence. Further, another 33.3% of the malaria vector were resistant to insecticides but lack any known kdr alleles.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 August 2023 doi:10.20944/preprints202308.1388.v1

Figure 5. Association between insecticide resistance and kdr allele presence.

The investigated association between the sporozoite rate and number of mosquitoes is shown in Figure 6. However, there was no significant association between the sporozoite rate and the population dynamics of mosquitoes. However, the sporozoite to the population dynamics of the malarial mosquitoes' ratio is in line with the random sampling de-facto distribution Hence, this could be explained that increase in the number of mosquitoes in a georeferenced MIS location, do not determine the increase in the sporozoite rate, and vice versa.

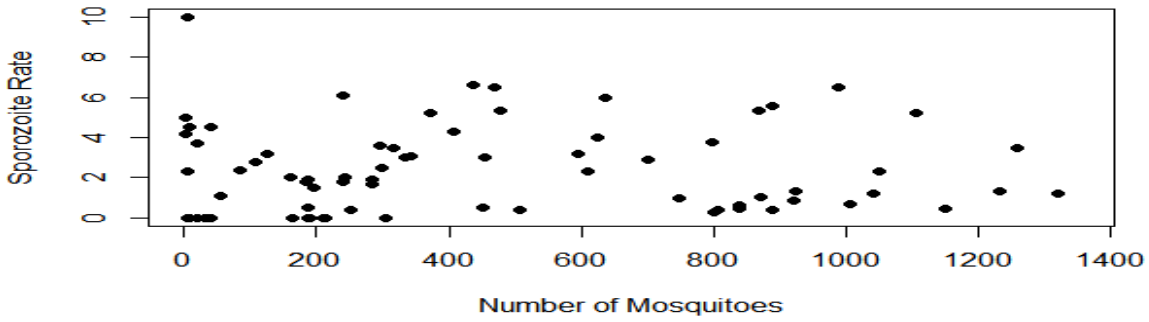


Figure 6. Association between the Sporozoite rate and the number of Malarial Mosquitoes.

4. Discussion

This work developed a computationally efficient novel E_Apriori algorithm that was faster theoretically and empirically in terms of its running time compared to existing Apriori algorithms. Further, it was simple to implement, cost-effective and specifically useful to generate accurate and graphically interpretable associative patterns on high-dimensional clinical and biological MIS data for the control of malaria.

This study found that there was no significant association between the wealth indices of the individuals and their respective anemia levels, which contradicted the existing hypothetical research inference [10,13,21,24,25]. Hence, this study suggested that confirmed malaria disease cases expressed through an anemic symptom, is not significantly associated with low-income individuals.

The findings from this study further contradicted the existing clinical hypothetical research inference, that most anemia symptomatic patients are expected to successfully accept malaria treatment [10,13,24]. However, this study found that there were other individuals, with no anemic symptom but were successfully treated of the malaria disease. Hence, this study hypothetically inferred that the malaria disease in an individual might not necessarily be expressed through an anemic symptom, which is useful for clinicians during malaria case evaluation.

This study suggests that there are other factors responsible for the insecticide resistance of the malarial mosquito asides kdr alleles present in the mosquito genetic make-up. This was because the study found a weak positive association between insecticide resistance and kdr alleles. Hence, the flight behavior, specific specie-strain of the malarial mosquitoes, which might be other contributing factors to its insecticide resistance capability, are recommended research niches for computational and experimental biologists.

This study further established a novel epidemiological trend that the sporozoite rate is not dependent on the population dynamics of the malarial vector. Hence, this suggests that the number of mosquitoes in a region do not determine the endemicity of the malaria disease in that region. Hence, this study recommended that malaria control-program formulators and clinicians should adopt a country-specific intervention and treatment approach, respectively towards the control of malaria.

Data Availability Statement: The Nigerian geo-referenced MIS 2015 data used for this study is downloadable on https://www.dhsprogram.com/data/. The supplementary literature-sourced Nigerian MIS data used for this study is downloadable on https://doi.org/10.1371/journal.pone.0028347.s001.

Acknowledgements: The author thank Prof. Jason Rasgon of the Department of Entomology, Center for Infectious Disease Dynamics, Pennsylvania State University, United States of America; for hosting him at his Lab during the period of carrying out this study between December 3, 2019.

Conflicts of Interest: None declared.

References

1. Karnad DR, Nor MB, Richards GA, Baker T, Amin P. Intensive care in severe malaria: report from the task force on tropical diseases by the World Federation of Societies of Intensive and Critical Care Medicine. Journal of critical care. 2018; 43:356-60. pmid: 29132978.
2. Gyapong J, Boatin B. Neglected tropical diseases-sub-Saharan Africa. Neglected Tropical Diseases. 2016; 87-112.
3. Onyishi GC, Aguzie IO, Nwani CD, Obiezue RN, Okoye IC. Malaria-Vector Dynamics in a Tropical Urban Metropolis, Nigeria. Pakistan Journal of Zoology. 2018; 50(3): 1035.
4. Parker M. Neglected Tropical Diseases. The International Encyclopedia of Anthropology. 2018; 1-2.
5. van Eijk AM, Hill J, Noor AM, Snow RW, ter Kuile FO. Prevalence of malaria infection in pregnant women compared with children for tracking malaria transmission in sub-Saharan Africa: a systematic review and meta-analysis. The Lancet Global Health. 2015; 3(10):e617-28. pmid: 26296450.
6. Damen JG., Daminabo VM. Prevalence of Malaria Parasitaemia in Pregnant Women WHO Attended General Hospital Shendam, Plateau State, Nigeria. Nature and Science. 2017; 15(1):10-17.
7. Rogerson SJ. Management of malaria in pregnancy. The Indian journal of medical research. 2017; 146(3):328. pmid: 29355138.
8. Gunda R, Chimbari MJ, Shamu S, Sartorius B, Mukaratirwa S. Malaria incidence trends and their association with climatic variables in rural Gwanda, Zimbabwe, 2005–2015. Malaria Journal. 2017; 16(1):393. pmid: 28964255.
9. Aregbeshola BS, Khan SM. Impact of health facilities on malaria control interventions among children under five years of age and pregnant women in Nigeria. South East Asia. Journal of Public Health. 2017; 7(1):35-41.
10. Simbauranga RH, Kamugisha E, Hokororo A, Kidenya BR, Makani J. Prevalence and factors associated with severe anaemia amongst under-five children hospitalized at Bugando Medical Centre, Mwanza, Tanzania. BMC hematology. 2015; 15(1):13. pmid: 26464799.
11. Sherer ML, Posillico CK, Schwarz JM. The psychoneuroimmunology of pregnancy. Frontiers in neuroendocrinology. 2018; 51:25-35. pmid: 29110974.
12. Willilo RA. Pregnant women and infants as sentinel populations to monitor prevalence of malaria: results of pilot study in Lake Zone of Tanzania. Malaria journal. 2016; 15(1):392. pmid: 27473039.
13. Morakinyo OM, Balogun FM, Fagbamigbe AF. Housing type and risk of malaria among under-five children in Nigeria: evidence from the malaria indicator survey. Malaria journal. 2018; 17(1):311. pmid: 30153834.
14. Dube B, Mberikunashe J, Dhliwayo P, Tangwena A, Shambira G, Chimusoro A, Madinga M, Gambinga B. How far is the journey before malaria is knocked out in Zimbabwe: results of the malaria indicator survey 2016. Malaria journal. 2019; 18(1):171. pmid: 31189478.
15. Builder MI, Anzaku SA, Joseph SO. Effectiveness of intermittent preventive treatment in pregnancy with sulphadoxine-pyrimethamine against malaria in Northern Nigeria. International Journal of Recent Scientific Research. 2019; 10(5): 32295-32299.
16. Okethwangu D, Opigo J, Atugonza S, Kizza CT, Nabatanzi M, Biribawa C, Kyabayinze D, Ario AR. Factors associated with uptake of optimal doses of intermittent preventive treatment for malaria among pregnant women in Uganda: analysis of data from the Uganda Demographic and Health Survey, 2016. Malaria journal. 2019; 18(1):250. pmid: 31349829.
17. Adeola A, Ncongwane K, Abiodun G, Makgoale T, Rautenbach H, Botai J, Adisa O, Botai C. Rainfall Trends and Malaria Occurrences in Limpopo Province, South Africa. International Journal of Environmental Research and Public Health. 2019; 16(24):5156. pmid: 31861127.
18. Shimaponda-Mataa NM, Tembo-Mwase E, Gebreslasie M, Achia TN, Mukaratirwa S. Modelling the influence of temperature and rainfall on malaria incidence in four endemic provinces of Zambia using semiparametric Poisson regression. Acta tropica. 2017; 166: 81-91. pmid: 27829141.
19. Okunlola OA, Oyeyemi OT. Spatio-temporal analysis of association between incidence of malaria and environmental predictors of malaria transmission in Nigeria. Scientific reports. 2019; 9(1):1-11.

20. Rouamba T, Nakanabo-Diallo S, Derra K, Rouamba E, Kazienga A, Inoue Y, Ouedraogo EK, Waongo M, Dieng S, Guindo A, Ouédraogo B. Socioeconomic and environmental factors associated with malaria hotspots in the Nanoro demographic surveillance area, Burkina Faso. BMC public health. 2019; 19(1):249.
21. Yunta C, Hemmings K, Stevenson B, Koekemoer LL, Matambo T, Pignatelli P, Voice M, Nász S, Paine MJ. Cross-resistance profiles of malaria mosquito P450s associated with pyrethroid resistance against WHO insecticides. Pesticide biochemistry and physiology. 2019; 161:61-67.
22. Thiaw O, Doucoure S, Sougoufara S, Bouganali C, Konate L, Diagne N, Faye O, Sokhna C. Investigating insecticide resistance and knock-down resistance (kdr) mutation in Dielmo, Senegal, an area under long lasting insecticidal-treated nets universal coverage for 10 years. Malaria journal. 2018; 17(1):123. pmid: 29566682.
23. Lynd A, Oruni A, van't Hof AE, Morgan JC, Naego LB, Pipini D, O'Kines KA, Bobanga TL, Donnelly MJ, Weetman D. Insecticide resistance in Anopheles gambiae from the northern Democratic Republic of Congo, with extreme knockdown resistance (kdr) mutation frequencies revealed by a new diagnostic assay. Malaria journal. 2018; 17(1):1-8. pmid: 30400885.
24. White NJ. Anaemia and malaria. Malaria journal. 2018; 17(1):371. pmid: 30340592.
25. Jasani S. Using a one health approach can foster collaboration through transdisciplinary teaching. Medical teacher. 2019; 41(7):839-841. pmid: 30071742.
26. Burkot TR, Bugoro H, Apairamo A, Cooper RD, Echeverry DF, Odabasi D, Beebe NW, Makuru V, Xiao H, Davidson JR, Deason NA. Spatial-temporal heterogeneity in malaria receptivity is best estimated by vector biting rates in areas nearing elimination. Parasites & vectors. 2018; 11(1):606. pmid: 30482239.
27. Getachew D, Gebre-Michael T, Balkew M, Tekie H. Species composition, blood meal hosts and Plasmodium infection rates of Anopheles mosquitoes in Ghibe River Basin, southwestern Ethiopia. Parasites & vectors. 2019; 12:257. pmid: 31122286.
28. Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. VLDB'94: Proceedings of the 20th International Conference on Very Large Data Bases. San Francisco, CA, USA. 1994:487-499.
29. Ai D, Pan H, Li X, Gao Y, He D. Association rule mining algorithms on high-dimensional datasets. Artificial Life and Robotics. 2018; 23(3):420-427.
30. Zhang C, Chen Y, Yang J, Yin Z. An association rule based approach to reducing visual clutter in parallel sets. Visual Informatics. 2019; 3(1):48-57.
31. Hegland M. The Apriori algorithm–a tutorial. Mathematics and computation in imaging science and information processing. 2007; 209-262.
32. Baffour KA, Osei-Bonsu C, Adekoya AF. A modified apriori algorithm for fast and accurate generation of frequent item sets. International Journal of Scientific & Technology Research. 2017; 6(8):169-173.
33. Huang L, Chen H, Wang X, Chen G. A fast algorithm for mining association rules. Journal of Computer Science and Technology. 2000; 15(1):619-624.
34. Yu H, Wen J, Wang H, Jun L. An improved Apriori algorithm based on the Boolean matrix and Hadoop. Procedia Engineering. 2011; 15(1):1827-1831.
35. Niu K, Jiao H, Gao Z, Chen C, Zhang H. A developed apriori algorithm based on frequent matrix. Proceedings of the 5th international conference on bioinformatics and computational biology. 2017; 5(1):55-58.
36. Li J, Sun F, Hu X, Wei W. A multi-GPU implementation of apriori algorithm for mining association rules in medical data. ICIC Express Letters. 2015; 9(1):1303-1310.
37. Liu X, Zhao Y, Sun M. An improved Apriori algorithm based on an evolution-communication tissue-like P system with promoters and inhibitors. Discrete Dynamics in Nature and Society. 2017; 6978146:1-12.
38. Yuan X. An improved Apriori algorithm for mining association rules. In AIP conference proceedings. 2017; 1820: 080005.
39. Okorie PN, McKenzie FE, Ademowo OG, Bockarie M, Kelly-Hope L. Nigeria *Anopheles* vector database: an overview of 100 years' research. PLoS One, 2011; 6(12): 1-16. pmid: 22162764.
40. Sunderland KM, Beaton D, Fraser J, Kwan D, McLaughlin PM, Montero-Odasso M, Peltsch AJ, Pieruccini-Faria F, Sahlas DJ, Swartz RH, Strother SC. The utility of multivariate outlier detection techniques for data quality evaluation in large studies: an application within the ONDRI project. BMC medical research methodology. 2019; 19(1):102. pmid: 31092212.