

## **MINING OF DRUGS REACTIONS IN PATIENTS WITH MALARIA USING APRIORI ALGORITHM**

*Adetunji A.B<sup>1</sup>, Alo O.O<sup>2</sup> & Popoola H.O<sup>3</sup>*

*<sup>1,3</sup>Research Scholar, Department of Computer Science, Ladoke Akintola University of Technology, Ogbomoso, Nigeria*

*<sup>2</sup>Research Scholar, Department of Information Systems, Ladoke Akintola University of Technology, Ogbomoso, Nigeria*

### **ABSTRACT**

*Data mining refers to the entire process of extracting useful and novel patterns or models from large data sets. With the widespread use of medical information systems that include databases, which have recently featured explosive growth in their sizes, physicians and medical researchers are faced with a problem of making use of the stored data. Data mining can be used to help predict future patient behavior and to improve treatment programs. By identifying high-risk patients, clinicians can better manage the care of patients today so that they do not become the problems of tomorrow.*

*One of the most dreaded diseases in Nigeria today is Malaria. Lots of drug has been discovered for this but it is noticed that most of the drugs are not effective in everyone. When a new drug is introduced unexpected drug reaction go unnoticed until large numbers of cases are reported by the diagnosed patients. Therefore, in exploring the capability of data mining so that the drug prescribed by the doctor is more efficient and of low risk of reactions to patients, this project was embarked upon.*

*Drug reaction can occur during treatment with pharmaceutical products. It can result in unnecessary and often fatal harm to patients. Several factors are responsible for the reaction of drug such as the patient's age, sex, blood group and genotype. Every anti-malarial has side effect and their rate of severity depends on these factors. Hence this model was developed to mine reaction of drugs in patients. The plots showed that different anti-malarial has different local support and also have different level of reaction compared to each other. From the model generated from the mined data set of the university health center using Apriori algorithm, it was recommended that, patient already having any symptom that is the same with the drug reaction that has exceeded the confidence value should not be given such drug.*

**KEYWORDS:** *Apriori Algorithm, Abnormal Class, Data Mining, Drugs Reaction, Local Support, Mysql, Php.*

---

### **Article History**

**Received: 11 May 2021 | Revised: 31 May 2021 | Accepted: 07 Jun 2021**

---

### **INTRODUCTION**

Data mining is the process of automatically discovering useful information in large data repositories. It is an integral part of Knowledge Discovery in Databases (KDD). (Frawly and Piatetsky-Shapiro, 1996). Knowledge Discovery in Database is a concept of the field of computer science that describes the process of automatically searching large volume of data for patterns that can be considered as knowledge about the data. Knowledge Discovery in Databases describes the overall

process of converting a series of transformation steps from data preprocessing to post-processing of data mining results. (Kantardzic, 2003).

Data mining is commonly used in wide range of profiling practices, such as marketing, surveillances, fraud detection and scientific discovery. Majority of areas related to medical services such as the prediction of effectiveness of surgical procedures, medical test, medication and the discovery of relationships of information is crucial for the health care organizations to stay competitive in today's complex, evolving environment. Data mining algorithms, such as reporting odds ratio (Van et al, 2002) or Multi-item Gamma Poisson Shrinker (Du Mouchel, 1999), have been used to generate and rank the difference drug- adverse effect associations or signals found in pharmacovigilance data.

Anti malarial drugs are the primary weapons to treat parasite infection, save lives, and curtail further transmission. There may be adverse events associated with the usage of these drugs. Commonly reported adverse events include cough, vomiting, itching. Li and Tian (2014) assessed adverse drug reaction in an oral antibiotics used in dermatological indications from the outpatient clinics. Li et al, (2005) also mined risk patterns in medical data. Study showed that adverse that adverse drug reaction accounted for some hospital admission. (Nivya et al, 2015) This also causes some death. (Hadi et al, 2017)

In this paper, Apriori algorithm is used on a health database containing records of malaria patients. The algorithm finds patterns with optimal relative risks. A pattern  $P$  is said to be frequent if its local support is greater than a set minimum local support. The local support of a pattern  $P$  is the ratio of records in the abnormal class. The abnormal class is a class of record containing a pattern under consideration. For example if a drug is selected as a parameter for the mining process, the abnormal class will contain all records for patients who used that drug. A frequent pattern whose relative risk is above a set minimum threshold is said to have an optimal relative risk. (Han et al., 2000, Li et al., 2005) These patterns can be used by medical practitioners for further medical research. Apriori algorithms have been used for analysis of consumer purchase patterns. (Suprianto et al, 2018).

## **MATERIALS AND METHODS**

### **Source of Data**

It is very essential, in order to obtain accurate data to use the correct methods of collecting statistical information. There are basically two main sources of data; these are the primary data and the secondary data. Primary data are obtained directly from the source while the secondary is the data collected by someone other than the user of the data. Secondary data analysis saves time that will otherwise be spent collating the database that might be unfeasible for any individual researchers to collect on their own. Secondary data generally have a pre established degree of validity and reliability which need not be examined by the researcher who is re – using the data. The data used in this project work is secondary, since it was collected from the kept records of the health centre.

### **Apriori Algorithm**

The Apriori algorithm developed by Agrawal and Srikantl (1994) is a great achievement in the history of mining association rules has referenced in Mingju and Sanguthevar, (2006). This technique uses the property that any subset of a large item-set must be a large item-set. Apriori uses bottom up approach, where frequent subsets are extended one item at a time (a step known as candidate generation) and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Apriori uses breadth first search and a tree structure to count candidate

item set efficiently. It generates candidate item set of length  $k$  from item set length  $k-1$ . Then it prunes the candidate which has an infrequent sub data. According to the downward closure lemma, the candidate set contains all frequent  $k - length$  item sets. After that it scans the transaction database to determine frequent item sets among the candidates.

The Algorithm

Procedure AprioriAlg()

Apriori ( $T, \epsilon$ )

```

→  $L_1 \leftarrow \{frequent\ 1 -\ itemsets\}$ 
→  $k \rightarrow 2$ 
→ while  $L_{k-1} \neq \phi$ 
→   →  $C_k \leftarrow \{a \cup \{b\} | a \in L_{k-1} \wedge b \notin a\} - \{c | \{s | s \subseteq c \wedge |s| = k-1\} \subseteq L_{k-1}\}$ 
→   →  $C_k := \text{apriori-gen}(L_{k-1})$  // New candidates
for all transactions  $t$  in the dataset  $T$  do
  for transaction  $t \in T$ 
    →   →  $C_k \leftarrow \{c | c \in C_k \wedge c \subseteq t\}$ 
for all candidates  $c \in C_k$  contained in  $t$  do
  →   → for candidates  $c \in C_k$ 
count  $[c] \leftarrow count[c] + 1$ 
→   →  $L_k \leftarrow \{c | c \in C_k \wedge count[c] \geq \epsilon\}$ 
→   →  $k \leftarrow k + 1$ 
 $L_k := \{c \in C_k | c.count \geq \text{min-support}\}$ 
return  $\bigcup_k L_k$ 

```

**ALGORITHM: Mining Drug Reactions in Patients with Malaria**

**Input**

Valid login and username, abnormal class ‘a’, minimum local support  $\delta$  in abnormal class a, and the minimum relative risk threshold  $\theta$ .

## Output

```

Optimal-Drug-Reactions-pattern-set-R.¶
Login(Username, Password)¶
If(Login() is successful) Then.¶
Select Drug¶
Else-re Login().¶
End-if¶
//Generats-Transaction-database¶
For-all-I-in-the-abnormal-class-do¶
→ TBD=(i,j)-where j=tid¶
//generate frequent patterns by pruning¶
For-distinct x in TBD-do¶
Get-local-support-x¶
→ If local-support-x ≥ δ¶
I:=(x,local-support-x)¶
End-if¶
End-for¶
//generate optimal-risk patterns.¶
For-distinct x in frequent-patterns-do¶
Get-relative-risk-x¶
If relative-risk-x ≥ θ¶
R:=(x,relative-risk)¶
End-if¶
End-for¶

```

The local support of a pattern  $P$  is the ratio of the number of records containing  $P$  to the total number of records in the abnormal class 'a'. A pattern is said to be frequent if its local support is greater than a set minimum local support  $\delta$ . A frequent pattern is said to have an optimal risk if its relative risk is greater than a set minimum threshold  $\theta$ . Over the years, frequent pattern mining has contributed a lot to the area of data mining. It helps in the extraction of previously unknown patterns which will be useful for making logical conclusions.

## RESULTS AND DISCUSSIONS

### The Patients Symptoms

The Patients Symptoms page enables the user enter the various patient symptoms into the database. The symptoms that could be added include: headache, Nausea, vomiting, itching, abdominal pain, blurred vision, hearing and diarrhea. After entering this information, then all the patients information's will be saved into the database. The Patient Symptoms Page is given in Figure 1.



Figure 1: The Patient Symptoms Page.

**The Abnormal Class**

The Abnormal Class Page shows the summary of all the information’s on the chosen drug from the user’s welcome page. The summary includes the total number of records in the database, number of records in the normal class and the number of records in the abnormal class. The “View Bar chart” in the abnormal class displays the records of all patients in the abnormal class in a bar chart format. These records are very essential in the mining process as it is one of the major implementation of the algorithm.

Abnormal Class for CAMOQUINE												
S/N	Age	Sex	BloodGroup	Genotype	Headache	Nausea	vomiting	Itching	Abdominal	Blurred	Hearing	Diarrhoea
1	18	M	O	AS	NO	YES	NO	YES	NO	NO	NO	NO
2	16	M	O	AA	NO	NO	NO	NO	NO	NO	YES	YES
3	21	M	O	AA	NO	NO	NO	NO	YES	NO	YES	NO
4	24	M	O	AA	NO	YES	NO	NO	NO	NO	YES	NO
5	21	M	O	AA	YES	YES	YES	NO	YES	NO	NO	NO
6	22	M	O	AA	YES	NO	NO	NO	YES	NO	NO	NO
7	24	M	O	AA	YES	NO	NO	NO	YES	NO	NO	NO
8	25	M	O	AS	YES	NO	NO	NO	YES	NO	NO	NO
9	19	F	AB	AS	YES	NO	YES	NO	YES	NO	NO	NO
10	23	F	AB	AS	YES	NO	YES	NO	YES	NO	NO	NO
11	16	F	O	AS	YES	NO	YES	NO	YES	NO	NO	YES
12	17	F	O	AS	YES	NO	YES	YES	NO	NO	NO	YES

LOCAL SUPPORT FOR DRUG	
Reaction Type	Value
Headache	0.67
Nausea	0.25
Vomiting	0.42
Itching	0.17
Abdominal Pain	0.67
Blurred Vision	0.00
Hearing	0.25
Diarrhoea	0.25

FREQUENT DRUG REACTION FOR DRUG	
Reaction Type	Value
Headache	0.67
Abdominal Pain	0.67

RELATIVE RISK FOR DRUG	
Reaction Type	Value
Headache	1.16
Abdominal Pain	3.81

OPTIMAL RELATIVE RISK FOR THE SPECIFIED DRUG	
Reaction Type	Value
Headache	1.16
Abdominal Pain	3.81
Hearing	1.04

Figure 2: The Abnormal Class Page for Camoquine.

The local support of a pattern  $P$  is the ratio of the number of records containing  $P$  to the total number of records in the abnormal class ‘a’. According to the figure 2, the local support for CAMOQUINE was given different symptoms. For example, the local support for CAMOQUINE on headache is 0.67, the local support on itching is 0.17 and the local support for blurred vision is 0.00. This means that CAMOQUINE is likely to cause headache since it has a higher value which is 0.67, CAMOQUINE is likely to cause itching in some patient since its value is lower (0.17)and CAMOQUINE will not cause blurred vision since the local support for blurred vision is 0.00.

The frequent drug reaction for a particular drug pattern is the number of records  $P$  whose local support is greater than the level of confidence. The level of confidence is the level in which the drug reaction can be said to be risky or dangerous if it is exceeded.

The relative risk of a drug is the risk of the drug to a reaction in relative to other the other drugs in the database. All the drugs in the database are considered when calculating the relative risk of a drug and only the reactions that could occur with the drug are considered.

In figure 2, the relative risk for CAMOQUINE is 1.16 for headache and 3.81 for abdominal pain. From the local support table, CAMOQUINE has a local support value of 0.67 for both the headache and the abdominal pain. Comparing this with the relative risk shows that CAMOQUINE causes headache in relative to other drugs with a value of 1.16 and that CAMOQUINE causes abdominal pain in relative to other drugs with a value of 3.81. This means that there are other drugs in the database that also causes headache in the database but there are only a few drugs that cause abdominal pain with respect to CAMOQUINE.

The optimal relative risk is derived from the relative risk of the drug. This shows the value of the optimum relative risk that could be caused by any drug. It makes it easier for the user to view the relative risk that could be caused by a drug.

Figure 3 shows the local support of Camoquine in a bar chart format. The Various symptoms of the patients in the database in the x-axis and the local support for each drug in the yaxis

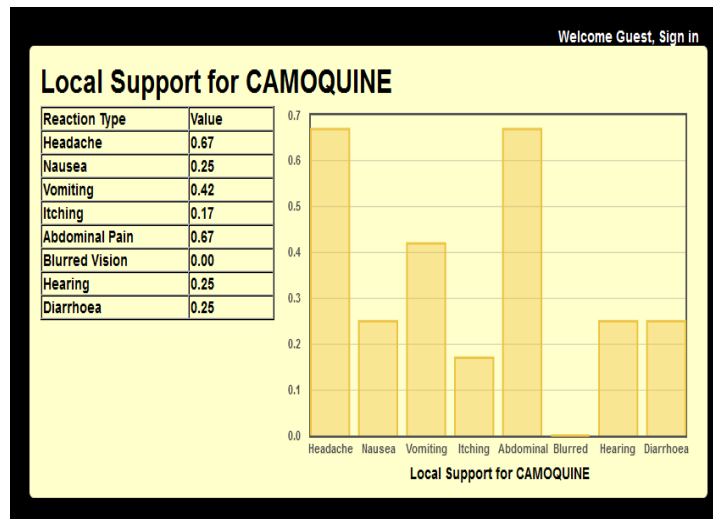


Figure 3: The Graph of Local Support for Camoquine.

**Compare Drugs**

Figure 4, the Compare Drugs Page, is the page shows the user the local support for all the drugs in the database, so that the user will be able to view all the local support that each drug has at a glance. This will enable the user to compare each drug with the total type of reactions that could accompany it.

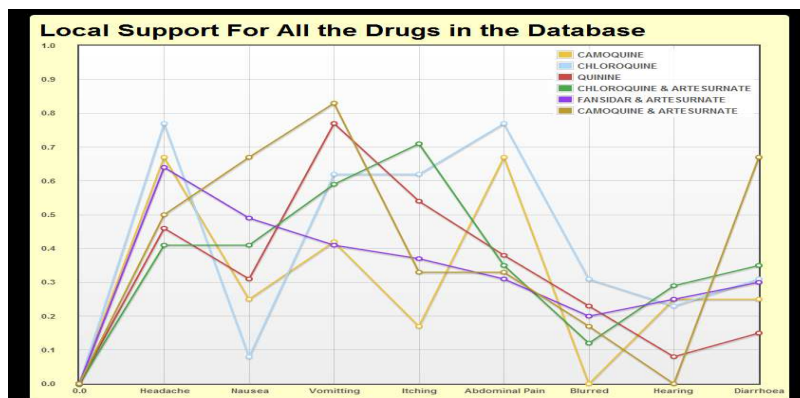


Figure 4: A Graph Showing the Local Support of all the Drugs in the Database.

From the graph, one would be able to conclude the reaction that could be caused a particular drug. CHLOROQUINE has the highest local support for headache, abdominal pain and blurred vision. CAMOQUINE AND ARTESURNATE has the highest local support for nausea and vomiting. CHLOROQUINE AND ARTESURNATE has the highest local support for itching, hearing and diarrhea.

With the result generated from the database, it shows that if a patient is having nausea or vomiting, then CAMOQUINE AND ARTESUNATE should not be prescribed for the patient since CAMOQUINE AND ARTESUNATE has the highest local support for vomiting and nausea. Likewise, if a patient is complaining of itching, hearing defect or diarrhea, then CHOLOROQUINE AND ARTESUNATE should not be prescribed for the patient since CHLOROQUINE AND ARTESUNATE has the highest local support for itching, hearing defect and diarrhea.

## CONCLUSIONS

As malaria is a dreaded disease in Nigeria today and very rampant in other community, we have been able to model that the anti malarial that are been used are effective and also have different level of reaction on students depending on their blood group, genotype, age and sex

A web based application using PHP and MYSQL has been used to implement Apriori algorithm for mining of drug reaction in patients with malaria using the university health center data repository to prove the effectiveness of data mining in the health sector.

The Apriori algorithm is used to find the patterns of drug reactions of patients in the database with the optimal relative risks. Apriori algorithm uses a minimum support value as the main constraint to determine whether a set of items is frequent. Frequent occurrences of items with each other are mined by Apriori to discover relationship among items and create a pattern. The pattern is checked against the minimum local support to know if it is greater than a set of minimum local support. For a particular reaction, if its generated pattern is greater than a set of minimum local support of a drug, then the drug is said to be frequent and could cause that reaction.

In conclusion, any anti – malarial which reaction is greater than the local support (0.5) is said to be a frequent reaction to patients. Such drug must not be prescribed to patient except for a special purpose. Optimal drug reactions should be mined on a regular basis to ensure improvements in the health care sector. After the administration of a drug, the healthcare organization should monitor the patients in order to take note of any other reactions different from those presumed to occur. Some drug reactions could be more severe than expected; patients should be given another drug alongside the anti malarial drug to help subside some adverse drug reactions.

## REFERENCES

1. Agrawal R. and Srikant R. (1994): "Fast algorithms for mining association rules in large data bases. In *Proceedings of the 20<sup>th</sup> International Conference on Very large databases, Santiago, Chile*
2. DuMouchel W.(1999): *Bayesian data mining in large frequency tables, with an application to*
3. *The FDA spontaneous reporting system. Am Stat; 53: 177–90.*
4. Frawly W.J. and Piatetsky-Shapiro (1996): *Knowledge Discovery in Databases: An Overview. The AAAI/MIT Press, Menlo Park, C.A.*

5. Hadi M.A., Neoh C.F., Zin R.M., Elrggal M.EE. and Cheema E.(2017): *Pharmacovigilance: Pharmacists' perceptive on spontaneous adverse drug reaction reporting. Integrated Pharmacy Res Pract.*; 6 91–8
6. Heon Gyu Lee, Ki Yong Noh, Keun Ho Ryu (2007): "Mining Bio signal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV," *LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining*, pp. 56–66.
7. Li J., A.W. Chee Fu, H. He, J. Chen, H. Jin, D. McAullay, G. Williams, R. Sparks, and C. Kelman (2005): Mining risk patterns in medical data, in: R. Grossman, R. J. Bayardo, and K. P. Bennet, eds, *Proceedings of the Eleventh ACM SIGKDD International conference on Knowledge Discovery and Data mining*, pp. 770–775.
8. Nivya K, Sri Sai Kiran V, Rago N, Jayaprakash B, Sonal Sekhar M.(2015): Systemic review on drug related hospital admissions – a pubmed based search. *Saudi Pharmaceut J.*;23(1):1–8
9. Niti Guru, Anil Dahiya, Navin Rajpal (2007): "Decision Support System for Heart disease diagnosis Using Neural Network", *Delhi Business Review*, Vol. 8, No. 1.
10. Van Puijenbroek EP, Bate A, Leufkens and HG, et al (2002): A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse Drug reactions. *Pharmacepidemiol Drug Saf*; 11:3–10.
11. Forgey (1965): *Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classification*". In: *Biometrics*
12. Suprianto Panjaitan, Sulindawaty, Mohammed Amin, Sri Lindawati, Ronal Watrianthos,
13. Hengki Tamando Shotang and Bosker Sinaga (2019): *Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns: Journal of Physics Conference Series*. 1255 012057.