American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-6, Issue-3, pp-25-27
www.ajer.org
Research Paper Open Access

Clustering dispersion of malaria disease in Saravan of Iran by data mining analysis

Fatemeh Hamzavi[1]
[1] *Department of Entomology, Faculty of Agriculture, Higher Educational Complex of Saravan, P. O. Box.9951634145 Saravan, Iran.*

ABSTRACT: *Malaria is a vector borne parasitic diseases in tropical and semi-tropical regions that transfer by anopheles mosquito to human. Epidemic of malaria disease depended on population of mosquito anopheles and the number of people that have malaria and existence of plasmodium in their bodies. Density of people that have malaria is index for assessment activity of anopheles mosquito. To describe the demographic of malaria in Sistan and Baluchistan Province, the data of malaria cases, collected by the disease management department of the health in Saravan, Iran from March 2010 to March 2017. For each patient, a form question was completed. The age, sex, nationality, time (year), job and type of parasite are attributes that they used for modeling. Result of Dunn's index for K-Means techniques shows k=4 is the maximum value and therefore, the optimal number of* CLUSTER *is 4 for k-means clustering algorithm. This paper have forecast of malaria patients, and proper information for have good program for control of this disease and anopheles mosquito. Malaria daisies most happen in March after raining, when the people go to the nature and village for vacation without enough covering. It not possible to eradication malaria in Saravan of Iran, because the people go out of border of Afghanistan and Pakistan and transfer malaria.*

Keywords–*Malaria, Anopheles, Plasmodium, mosquito*

I. INTRODUCTION

Malaria is transmitted among humans by female mosquitoes of the genus *Anopheles*. Female mosquitoes take blood meals to carry out egg production, and such blood meals are the link between the human and the mosquito hosts in the parasite life cycle. Like all mosquitoes, anopheles goes through four stages in their life cycle: egg, larva, pupa, and adult. The first three stages are aquatic and last 5-14 days, depending on the species and the ambient temperature. Malaria mosquito lives in The south-eastern part of Iran, consisting of Sistan and Baluchistan Province, Hormozgan Province and the tropical part of Kerman Province with a combined population of approximately three million, is considered to be a 'refractory malaria region. Several factors such as climate diversity in the region have important role in density of mosquito population (1). We should remind that Sistan and Baluchistan is the largest Province in Iran. The districts of the province are Iranshahr, Chabahar, Khash, Zahedan, Saravan, Nikshahr, and Sarbaz. Being geographically adjacent to *Plasmodium falciparum*endemic regions of Afghanistan and Pakistan (2, 3).This province is bordered from the south with Oman Sea, Indian Ocean; and has a subtropical climate. Becauseof appropriate temperature and humidity, complete cycle of malaria mosquito is optimal from April to October in some parts of the province.Climate changes were shown to have a relationship with an increase in the anopheles mosquito incidence in the northwest frontier province of Pakistan, geographically next to the field of our study (2). Through numerous mechanisms, this cyclone might lead to increased incidence rates. Attenuation of transmission rate due to the developing of stagnant waters and subsequently higher mosquito density in the air, impairing malaria control measures from diagnosis to treatment in inaccessible areas, and subsequently the vulnerability might be proposed as some contributory mechanisms. We aimed in this paper study the malaria patient and clustering patient with character for more management and forecasting.

II. Materials and Methods

To describe the demographic of malaria in Sistan and Baluchistan Province, the data of malaria cases, collected by the disease management department of the health in Saravan, Iran from March 2010 to March 2017. For each patient, a form question was completed. Questions were about age, sex, nationality, time (year) , job and type of

parasite. According the aim of this research, K-Means clustering method was selected for identifying the cluster of this disease. Clustering methods help discover groups of data records with similar values or patterns. These techniques are used in marketing (customer segmentation) and other business applications (4). K-means clustering is a relatively quick method for exploring clusters in data (4) and can be used when we don't know what distinct groups are at the beginning (5). The k-means method aim is to minimize the sum of squared distances between all points and the cluster center (6). In this algorithm user sets the number of clusters (k) and each data record is then assigned to the nearest of the k clusters. This procedure is typically run several times because the user must set k and algorithm runs for each k. Therefore, this algorithm is an iterative algorithm (7). The k-means clustering algorithm ran on under review dataset with K=2, 3, 4, 5, 6, 7 and 8 values for number of clusters. The dataset variables are age, sex, nationality, time (year), job and type of parasite. The Dunn's index was used for calculating the optimal number of clusters. Dunn's Validity Index attempts to identify those cluster sets that are compact and well separated [7]. The aim of Dunn's index is to maximize the inner cluster distance and minimize the outer cluster distance. If obtained value for Dunn's index is large then it is better.

III. Results

Calculation of Dunn's index for k= 2, 3, 4, 5, 6, 7 and 8, shows the obtained value for k=4 is larger. Fig 1 shows k=4 is the maximum value and therefore, the optimal number of cluster is 4 for k-means clustering algorithm. Table 1 describes 4 clusters that obtained by k-Means algorithms.

The **headings** and **subheadings**, starting with "**1. Introduction**", appear in upper and lower case letters and should be **set in bold and aligne**

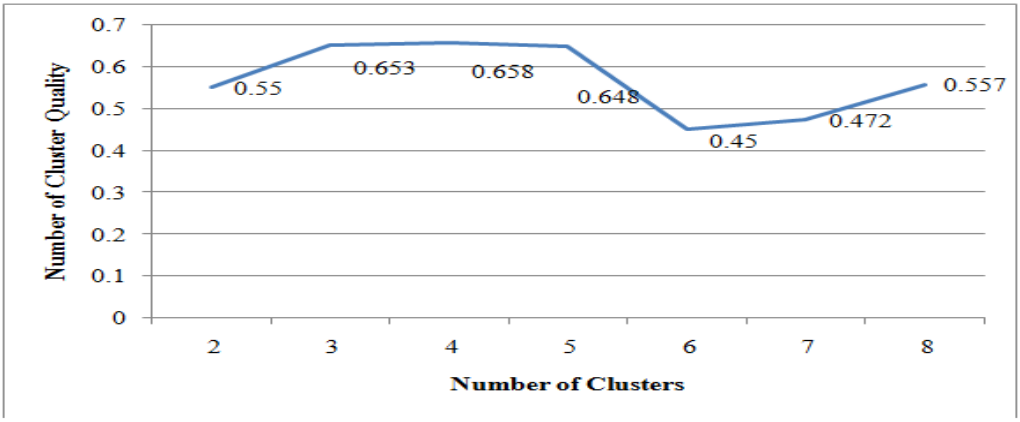

Figure 1: Number of Cluster Quality for K=2, 3, 4, 5, 6, 7 and 8 in K-means Algorithms

Table 1: Result of K-Means Algorithms for K=4

Attribute	Cluster 1	Cluster 2	Cluster 3	Cluster 4
age	26.377±15.907	28.058±19.006	26.677±14.711	27.677±14.058
Sex				
male	100	0	100	97.92
female	0	100	0	2.08
Date				
2010	3.77	5.77	1.59	4.17
2011	9.43	1.92	4.76	14.58
2012	5.66	11.54	1.59	25
2013	7.55	5.77	7.14	43.75
2014	43.4	28.85	19.05	12.50
2015	5.66	19.23	32.54	0
2016	13.21	13.46	25.40	0
2017	11.32	13.46	7.94	0
Job				
baby	11.32	19.23	7.14	4.17
driver	9.43	0	14.29	0
employment	5.66	0	1.59	4.17
farmer	0	0	1.59	2.08
Self-employed	35.85	1.92	30.16	79.17
Housewife	1.89	69.23	0	00
elderly	1.89	1.92	1.59	6.25

student	7.55	7.69	8.73	2.08
unemployment	0	0	0.79	2.08
worker	26.42	0	34.30	85.42
Nationality				
Iranian	67.92	75	69.05	4.17
Afghan	11.32	5.77	11.11	10.42
Pakistn	20.75	19.23	19.84	0
Type of malaria				
Falciparum	92.45	17.31	0	0
Vivax	0	69.23	100	27.08
Mix	7.55	13.46	0	72.92

Cluster 1 contains 53 samples patient have malaria 36.377 ± 15.907 value for average of age. They are men that have self-employment, 67% Iranian. 92% of patient had P. falciparum in 2014 year. Cluster 2 contains 52 samples patient have malaria 28.058 ± 19.006 value for average of age. They are women that have housewife job, 67% Iranian. 69.23% of patient had P. vivax in 2014-2017 year. Cluster 3 contains 126 samples patient have malaria 26.667 ± 14.711 value for average of age. They are men that have worker or self-employment job, 69% Iranian, all of patient had P. vivax in 2014-2015 year. Cluster 4 contains 48 samples patient have malaria 27.667 ± 14.058 value for average of age. 97.92% of them are men that have 79% self-employment job, 85% Iranian. 73% of patient had mix malaria in 2013 year. The same results were formerly reported. Rafi et al (8) concluded Malaria was most common in 15-44 year old people, rural areas, and males. Plasmodium vivax was the cause of disease in (83.8%) of patients, Plasmodium Falciparum in (13.4%) of patients and mixed species in (2.8%) of patients. And they showed that 15-20% of the total malaria patients had Afghani or Pakistani nationality. Burdens of malaria in Afghanistan and Pakistan induced by their specific sociopolitical challenges (9, 10) are hazardous factors that affect eradication of malaria in Iran. Similar findings were reported by a spatial study performed in this province formerly. That spatial modeling revealed that humidity, temperature, and altitude were positively correlated with the malaria risk. At feature programs included insecticide spraying, entomology survey and environment management. Elimination program for national malaria needed to control program for anopheles mosquito. Like this paper Ashori and Taheri (11) Using Clustering Methods for Identifying Blood Donors and have forecast some for feature.

IV. CONCLUSION

This paper have forecast of malaria patients and anopheles activity, and proper information for have good program for control of this disease and anopheles mosquito. This study showed the most patient were in April and May every year and percent of patient more rose up especially the weather was rainy in March and February and October. This condition increased anopheles population, Consequence increased malaria. At April in Iran people go to vacation in village and nature near the temporary water without suitable covering. And mosquito anopheles vectored the malaria. Every year have seen some people of Afghan and Pakistan nationality, migrated to Iran. There is one of the most factors prevented of eradication malaria in Iran.

V. Acknowledgements

The author would like to offer particular thanks to Miss. M. Ashori, responsible for guide and data analyses.

REFERENCES

[1.] Salehi M, Mohammad K, Farahani MM, Zeraati H, Nourijelyani K,
[2.] Zayeri F. Spatial modeling of malaria incidence rates in Sistan and Baluchistan province, Islamic Republic of Iran. *Saudi Med J* 2008;**29**:1791-6.
[3.] Bouma MJ, Dye C, van der Kaay HJ. Falciparum malaria and climate change in the northwest frontier province of Pakistan. *Am J Trop*
[4.] *Med Hyg*1t996;**55**:131-7.
[5.] Dost AG, Muslim M. Malaria in Afghanistan. *Med Parazitol (Mosk)*2001;**(1)**:42-3.
[6.] SPSS Inc. Introduction to Clementine®. 2003.
[7.] SPSS Inc. Clementine® 8.0 User's Guide. 2003.
[8.] Ray, S., Turi, R.H. Determination of Number of Clusters in K-Means Clustering and Application in Color Image Segmentation. 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99), 1999, PP: 137-143.
[9.] Ansari, Z., Azeem, M.F., Ahmed, W., Babu, A.V. Quantitative Evaluation of Performance and Validity Indices for Clustering the Web Navigational Sessions. World of Computer Science and Information Technology Journal (WCSIT), Vol. 1, No. 5, 2011, PP:217-226.
[10.] Rafi SM, Rashid M, Shorey WU. Malaria survey of border area of Baluchistan adjacent to Iran.*Pak JHealth* 1957;**6**:233-42.
[11.] 9 .Kolaczinski J, Graham K, Fahim A, Brooker S, Rowland M. Malaria control in Afghanistan: progress and challenges. *Lancet* 2005;**365**:1506-12. [15850637] [doi:10.1016/S0140 6736(05)66423-9]
[12.] Rowland M, Rab MA, Freeman T, Durrani N, Rehman N. Afghan refugees and the temporal and spatial distribution of malaria in Pakistan. *SocSci Med* 2002;**55**:2061-72. 36(01).
[13.] Ashoori, M., Taheri , Z. 2013.Using Clustering Methods for Identifying Blood Donors Behavior.5TH Iranian conferences on electronica and electronica engineering.Islamic Azad University Gonabad Branch, August 20-21-22, 2013.