Malaria continues to pose a serious threat to public health in North- Western states of India. Sahibganj is highly endemic for Malaria predominately with Plasmodium falciparium infections. Despite continuous efforts by NVBDCP, a desirable level of control has not been achieved. The present study describes the application of self organizing maps , a data mining tool for prioritization of malaria endemic zones in this region.1 District hospital and 7 PHCs (Public Health Centers) were randomly selected from Sahibganj and 6 maliariometric parameters via Annual Blood Examination rate (ABER), Annual Parasite Incidence (API), Slide Positivity Rate (SPR), Annual Falciparum Incidence (AFI) and Slide Falciparum Rate (SFR) were considered which reflected the intensity of malaria transmission in this region. Self Organizing Maps based on neighborhood distance, which reflects about zones based on status of intensity of malaria epidemiology. Such maps would make it possible to target control measures at high-risk areas and greatly increase the cost efficiency of malaria control programmes.

© 2013 Universal Research Publications. All rights reserved

**Abstract**

Malaria continues to pose a serious threat to public health in North- Western states of India. Sahibganj is highly endemic for Malaria predominately with Plasmodium falciparium infections. Despite continuous efforts by NVBDCP, a desirable level of control has not been achieved. The present study describes the application of self organizing maps, a data mining tool for prioritization of malaria endemic zones in this region. 1 District hospital and 7 PHCs (Public Health Centers) were randomly selected from Sahibganj and 6 maliariometric parameters via Annual Blood Examination rate (ABER), Annual Parasite Incidence (API), Slide Positivity Rate (SPR), Annual Falciparum Incidence (AFI) and Slide Falciparum Rate (SFR) were considered which reflected the intensity of malaria transmission in this region. Self Organizing Maps based on neighborhood distance, which reflects about zones based on status of intensity of malaria epidemiology. Such maps would make it possible to target control measures at high-risk areas and greatly increase the cost efficiency of malaria control programmes.

**Introduction**

Malaria, the third leading cause of death attributable to an infectious disease worldwide, has plagued mankind for countless generations. Malaria remains a public health problem in 90 countries in the world and causes more than 300 million acute illnesses and at least one million deaths annually. Malaria has been a problem in India for centuries. Details of this disease can be found in the ancient Indian Medical Literature like the “Charaka Samhita”. Now The annual incidence of malaria in India estimates to nearly 1 to 1.5 million cases. In this region, efficient malaria transmission is maintained during most months of the year and slashes potential economic growth and thus is a major impediment to the overall development and progress of these areas.

Despite of several anti-malaria programmes being implemented under National Vector Borne Diseases Control Programme, this region has seen little tangible progress in alleviating the burden of malaria. Apparently, there are definite inadequacies that continue to dampen the spirit of public health specialists even during the halcyon days of malaria eradication.

On closer scrutiny, it was evident that, there being financial and technical constraints common to all states of India, operational difficulties are hampering the effective malaria control in the North-Western region. These very areas remain inaccessible owing to floods and poor road communication. The major reasons of perennial and persistent malaria transmission are predominance of Plasmodium falciparum, difficult terrain, congenial eco-climatic conditions, lack of proper implementation of control operations, ineffective communication between health researchers and policy makers. The problem of drug resistance exophilic and exophagic vector behaviour and high efficiency of vectors further aggravate the situation.

Due to these various factors encountered in the North-Western region, malaria continues to present health services with an immensely difficult and complex challenge. The highly focal nature of malaria requires targeting of interventions to specific regions and malaria control interventions must be preceded by the identification and prioritization of the most vulnerable. Hence, there is an imperative need for exploitation of advanced Information Technology tools which can prioritize the endemic zones to liberate the region from manacles of this pandemic. Information technology has been successfully exploited in different spheres of control of vector-borne diseases. Computer applications in database management and data mining have paved way for control of malaria and Filariasis and have proved to be a valuable tool in decision making in vector identification. Computer simulation models of vector borne diseases have been used to develop early warning system for the epidemics. Various expert systems have also been developed employing Artificial Intelligence (AI) for insect pest management and for forecasting of outbreak of vectors. In accordance with their important role
in easy visualization of complex epidemiological data, Self Organizing Maps (SOM) are customized and used in the present study to prioritize the malaria endemic zones for disease management in Sahibganj.

**Materials and Methods**

Study site: Sahibganj is the smallest District in Jharkhand state area-wise situated in The climate is medium and humid at the lower altitudes and in the valleys covered by swampy dense forest particularly in the eastern section, while it becomes exceedingly cold in winter month. Average temperature during the summer month’s ranges from 40°C to 45°C during monsoon. The district lies approximately between 24°42’ and 25°20’ north latitudes and between 87°25’ and 87°54’ east longitudes. This district covers an area of 1599 km².

The district can be divided into two natural divisions based on its geographical features. The first region consists Borio, Mandro, Barhait, Pathna and Taljhari blocks and it is a part of the *Dumani-koh* region. The hills and its slopes are covered with forests, once dense but scanty now. The second region consists Sahibganj, Rajmahal, Udhwa and Barharwa blocks. This plain region comprises the uplands, undulation along ridges and depressions. The Ganga, Gumani and Bansloi rivers flow through this region. Forested terrain and perennial streams are congenial for rapid multiplication and longevity of malaria vectors.

Population of state is estimated to be 1163489 according to 2012 estimation. The state has a major population of 40 scheduled tribes and numerous sub-tribes. Agriculture is the primary driver of the economy. Nearly 80% of the population is engaged in agriculture. The traditional method of agriculture is Jhumming, a kind of shifting cultivation. The main crops are rice, maize, millet, wheat and mustard.

**Methods**

Data Collection: Raw data was collected from the District malaria office Sahibganj, which consists of Epidemiological aspects of Malaria cases encountered in District Hospital & 7 randomly selected Public Health Centers belonging to 141 HSCs and 1819 villages districts of Sahibganj in 2011 and 2012. Raw data pertaining to malaria incidence was collected and standard malariometric parameters (ABER, API, SPR, SFR, AFI(Pf%)) were calculated based on this data to be used in this study.

**Parameters identified for SOM:** Parameters like Annual Blood Examination rate (ABER), Annual Parasite Incidence (MPI), Slide Positivity rate (SPR), Slide *Falciparum* rate (SFR) and Annual Falciparum Rate (AFR) and *Plasmidium falciparum* % were considered for this study. All the factors were given equal importance.

Data Analysis: Data mining – Self Organizing Maps In SOM, neurons compete with each other to earn the table-1

**Table 1: Parameters identified for SOM - Sahibganj**

<table>
<thead>
<tr>
<th>Sahibganj</th>
<th>Population</th>
<th>BS.Coll</th>
<th>Total +ve cases</th>
<th>Pf cases</th>
<th>ABER</th>
<th>API</th>
<th>SPR</th>
<th>SFR</th>
<th>Pf%</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1136220</td>
<td>101466</td>
<td>5941</td>
<td>6648</td>
<td>8.93</td>
<td>5.23</td>
<td>5.86</td>
<td>4.62</td>
<td>78.8</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1163489</td>
<td>103357</td>
<td>3784</td>
<td>3372</td>
<td>8.88</td>
<td>3.25</td>
<td>3.66</td>
<td>3.26</td>
<td>89.1</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2: Malariometric indices of Public Health centers in Sahibganj**

<table>
<thead>
<tr>
<th>Sl. No</th>
<th>Name of PHC</th>
<th>Year</th>
<th>Populatio n</th>
<th>BS.Coll</th>
<th>Total +ve cases</th>
<th>Pf cases</th>
<th>Increase/Decrease</th>
<th>ABER</th>
<th>API</th>
<th>SPR</th>
<th>SFR</th>
<th>Pf%</th>
<th>Deaths</th>
<th>% Increase/ Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sadar Hospital Sahibganj</td>
<td>2011</td>
<td>0</td>
<td>10906</td>
<td>975 769</td>
<td>-280</td>
<td>-322</td>
<td>0</td>
<td>0</td>
<td>8.94</td>
<td>7.05</td>
<td>78.9</td>
<td>1</td>
<td>-2</td>
</tr>
<tr>
<td>2</td>
<td>Sadar Block</td>
<td>2012</td>
<td>10397</td>
<td>136159</td>
<td>5962</td>
<td>124</td>
<td>100</td>
<td>-2588</td>
<td>-2159</td>
<td>4.05</td>
<td>0.89</td>
<td>5.19</td>
<td>1.77</td>
<td>80.6</td>
</tr>
<tr>
<td>3</td>
<td>Borio</td>
<td>2011</td>
<td>207977</td>
<td>14450</td>
<td>839 827</td>
<td>-145</td>
<td>-34</td>
<td>6.79</td>
<td>3.94</td>
<td>5.82</td>
<td>5.73</td>
<td>98.6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Barhait</td>
<td>2012</td>
<td>137636</td>
<td>20785</td>
<td>984 861</td>
<td>479</td>
<td>398</td>
<td>15.10</td>
<td>7.15</td>
<td>4.73</td>
<td>4.14</td>
<td>87.5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Pathna</td>
<td>2011</td>
<td>87046</td>
<td>12422</td>
<td>166 113</td>
<td>92</td>
<td>88</td>
<td>16.34</td>
<td>1.91</td>
<td>1.77</td>
<td>0.79</td>
<td>68.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Taljhari</td>
<td>2012</td>
<td>89135</td>
<td>15295</td>
<td>74 25</td>
<td>-98</td>
<td>-61</td>
<td>17.16</td>
<td>0.83</td>
<td>0.48</td>
<td>0.16</td>
<td>33.8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Rajmahal</td>
<td>2011</td>
<td>334147</td>
<td>19296</td>
<td>505 470</td>
<td>-157</td>
<td>57</td>
<td>7.52</td>
<td>1.52</td>
<td>2.62</td>
<td>2.44</td>
<td>93.1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>Malariometric indices of Public Health centers in Sahibganj</td>
<td>2012</td>
<td>30023</td>
<td>6490</td>
<td>346 326</td>
<td>5595</td>
<td>-4358</td>
<td>7.21</td>
<td>3.84</td>
<td>5.33</td>
<td>5.02</td>
<td>94.2</td>
<td>0</td>
<td>-4</td>
</tr>
</tbody>
</table>

of representing the input data. As a result, data analysis attribute space can be abstracted to a much smaller number of latent formula organized on a basis of a parameter identified. Via this way the structures embedded in the input data can be revealed which is placed in the input space and is spanned over the inputs distribution. Using a SOM network, it is possible to obtain input space represents closeness of the input data. Processing units in the SOM lattice is associated with weights of the same dimension of the input data. During the learning stage the weights of the units change their position and “move” towards the input points. This “movement” becomes slower and at the end of the learning stage, the network is “frozen” in the input space. In SOM summarizing input data API >2 that indicated that highly endemic area.

After the learning stage the inputs can be associated to the nearest network unit. When, the inputs can be associated to each cell on the parameter. One or more cell that clearly contains similar objects can be considered as a Graph on the parameter. These Graphs are generated during the learning phase without any other information.
Results
Normalized data is Graphed using SOM yielded (shown in Graph1). Unsupervised learning was done on the fly using the data using a learning constant of API in to PHC in equal remaining PHCs fluctuation data iterations following which the data got Graphed among Graphs based on the neighborhood distance.

Discussion
Graph (1): Though the Sahibganj under this Graph shows low to moderate incidence of malaria yet compare than 2011 and 2012 a low in API, SPR, SFR.
Graph (3): shows that, Sader Block low ABER and thus cuts a poor picture where because of the lack of appropriate surveillance, there is a possibility of underestimating the disease burden. There is an immediate need to investigate the causes hindering the proper survey and strengthening the HealthCare infrastructure. In spite of moderate SPR observed, some regions like Taljhari and Borio are showing very high Pf% and thus requires immediate attention to curb this deadly parasite from its transmission to other regions.

MAP OF SAHIBGANJ DISTRICT

Graph 1: SOM Graphs showing in Sahibganj Comparing 2011 & 2012
Graph (3): Only one PHC Barhait ABER low in comparative in 2011 & 2012 in this Graph and indicates moderate incidence of malaria. This region requires both drug administration and vector control measures at relatively less priority.
Graph (4): Low to moderate SPR and SFR observed the Pathana PHCs this Graph low falciparum infections indicates that drug administration and vector control operations can be done with less priority. Graph (2, 3): The PHCs in this Graph shows an alarming rate of high malarial indices. Taljhari PHCs shows very high level of all malarialometric parameters and reflects high endemicity. The PHCs Graphed in this Graph are showing predominance of P. falciparum and thus drugs particularly targeting drug resistant falciparum malaria should be administered in this region. Inspite of efforts and various control operations undertaken by NMEP, malaria is still deeply entrenched in these regions. These regions warrant further investigation and more focused efforts on active surveillance and require both drug administration and vector control measures at relatively high priority and need a radical overhaul in the way it tackles the disease. Graph (2, 4): 2 PHCs Graphed in this region shows moderate level of all malarialometric indices and clearly reflects high falciparum trend as indicated by SFR, AFI. Only still shows very high API, and thus needs malaria control interventions at.

Graph 2: SOM Graphs showing different endemicity levels of Public Health centers. (2011 & 2012)
API= Annual Parasite Incidence,

Graph 3: SOM Graphs showing different endemicity levels of Public Health centers. Comparing 2011 & 2012
ABER= Annual Blood Examination Rate,

Graph 4: SOM Graphs showing different endemicity levels of Public Health centers. Comparing 2011 & 2012
SPR= Slide Positive Rate,
high annual Parasite incidence (API) and moderate to high SPR but very low or almost negligible *P. falciparum* incidence is reflected by SFR and AFL1.3 PHCS namely Taljhari, Borio are demonstrating very high API This shows that despite highly satisfactory surveillance leading to high ABER malaria was still refractory to intervention measures and hence, there is a need to excogitate our strategy for control again.

**Conclusions**
The application of Data mining and artificial intelligence in Epidemiology is still in its infancy. In spite of numerous evidences of incorporation of artificial intelligence as an aid in data analysis of various epidemiological studies, medical entomologists are still unable to tap its potential in vector control except for data acquisition and storage. Information Technology in vector control operations has been extended to construction of Databases on different aspects of vector borne diseases and various forecasting systems based on computer simulation models. Application of Artificial Intelligence in combating vector borne diseases can give a completely new dimension to existing control programs. Artificial neural Networks such as Kohonen Maps have a natural propensity to learn—they learn how to solve problems from data as opposed to solving problems based on explicit problem specification. Self Organizing maps (SOM) are deemed as being highly effective as a sophisticated visualization tool for visualizing high dimensional complex data with inherent relationships between the various features comprising the data. These have been successfully exploited in Medical and Health Informatics in fields as varied as Medical image processing, disease diagnosis, gene prediction, gene sequence analysis, expression analysis structural recognition of protein families and drug designing and drug utilization. In recent past, SOMs have been employed for data exploration in major public health diseases like Diabetes, Glaucoma. In this paper, we have shown the use of Self Organizing Maps as valuable tool in prioritization of malaria endemic zones which will assist in decision making on the location and deployment of health care services and prioritization of intervention strategies. In areas like sahibganj which suffer from perennial malaria transmission, and where difficult terrain and geographical features are big hurdles in carrying out effective and timely vector control operations, SOM will be very effective in bridging the gap between policy makers and Health workers. Recognizing consistent foci of cases would permit control efforts to be directed at specific geographic areas, reducing costs and increasing effectiveness. In a country like ours where resources are scarce, reliable methods for the stratification of zones on basis of the prevalence or transmission intensity of malaria are urgently required. Such Graphing and data visualization tools are essential for assessing the austerity of the problem, and hence the resources needed to emulate malaria. This approach will serve as yardstick for assessing the progress of control and indicate which geographic areas should be prioritized, so that large amount of man power and resources can be saved. Because of underlying simplicity in data visualization, SOM will prove to be a powerful weapon in arsenal in fight against this dreaded disease. This strategy will play a crucial role in bridging research and control and it is quite likely that besides reducing the malaria burden, the entire public health system will benefit from such a strategy if adopted and extrapolated to other regions across the world for other vector borne diseases

**ACKNOWLEDGEMENTS:**
The authors are thankful to The State program me officer and District malaria officer for providing facilities during this work was carried out.

**References**
5. Kondrashin AV, Rooney W and Singh N. Dynamics of P. falciparum ratio - An indication of malaria resistance or a result of control measures? Indian J Malarial 1987; 24: 89-94

Source of support: Nil; Conflict of interest: None declared