RISK ZONE MODELLING AND EARLY WARNING SYSTEM FOR VISCERAL LEISHMANIASIS (KALA-AZAR) DISEASE IN BIHAR, INDIA USING REMOTE SENSING AND GIS

A. Jeyaram a, Shreekant Kesari b, Anju Bajpai a, Gouri sankar Bhunia b, YVN Krishna Murthy c

a Regional Remote Sensing Centre - East, National Remote Sensing Centre, Indian Space Research Organisation, Kolkata, India, ajeayaram2002@yahoo.co.in

b Department of Vector Biology and Control, Rajendra Memorial Research Institute of Medical Sciences (ICMR), Patna, India

c Dy. Director (RC), National Remote Sensing Centre, ISRO, Hyderabad, India

Commission VIII, WG VIII/2

KEY WORDS: Visceral Leishmaniasis, Kala-azar, vector density, Satellite data, Remote Sensing, GIS

ABSTRACT:

Visceral Leishmaniasis (VL) commonly known as Kala-azar is one of the most neglected tropical disease affecting approximately 200 million poorest populations ‘at risk’ in 109 districts of three endemic countries namely Bangladesh, India and Nepal at different levels. This tropical disease is caused by the protozoan parasite Leishmania donovani and transmitted by female Phlebotomus argentipes sand flies. The analysis of disease dynamics indicate the periodicity at seasonal and inter-annual temporal scale which forms the basis for development of advanced early warning system. Study area of highly endemic Vaishali district, Bihar, India has been taken for model development. A systematic study of geo-environmental parameters derived from satellite data in conjunction with ground intelligence enabled modelling of infectious disease and risk villages. High resolution Indian satellites data of IRS LISS IV (multi-spectral) and Cartosat-1 (Pan) have been used for studying environmentally risk parameters viz. peri-domestic vegetation, dwelling condition, wetland ecosystem, cropping pattern, Normalised Difference Vegetation Index (NDVI), detailed land use etc towards risk assessment. Univariate analysis of the relationship between vector density and various land cover categories and climatic variables suggested that all the variables are significantly correlated. Using the significantly correlated variables with vector density, a seasonal multivariate regression model has been carried out incorporating geo-environmental parameters, climate variables and seasonal time series disease parameters. Linear and non-linear models have been applied for periodicity and inter-annual temporal scale to predict Man-hour-density (MHD) and ‘out-of-fit’ data set used for validating the model with reasonable accuracy. To improve the MHD predictive approach, fuzzy model has also been incorporated in GIS environment combining spatial geo-environmental and climatic variables using fuzzy membership logic. Based on the perceived importance of the geo-environmental parameters assigned by epidemiology expert, combined fuzzy membership has been calculated. The combined fuzzy membership indicate the predictive measure of vector density in each village. A Y factor has been introduced to have increasing effect in the higher side and decreasing effect in the lower side which facilitated for prioritisation of the villages. This approach is not only to predict vector density but also to prioritise the villages for effective control measures. A software package for modelling the risk villages integrating multivariate regression and fuzzy membership analysis models have been developed to estimate MHD (vector density) as part of the early warning system.

1. INTRODUCTION

1.1 General

India has 52 kala-azar endemic districts in four states of India - Bihar (31 districts with 62.3 million population), Uttar Pradesh (11 districts with 50.0 million population), West Bengal (6 districts with 11.0 million population), and Jharkhand (4 districts with 6.7 million population). Each year, Bihar alone contributes 70-80 % of the kala-azar cases. In 1992, the highest number of cases were reported (77170) after which the incidence has declined gradually. However, from 2003 to 2007 there has been a steady increase in the annual number of reported cases. In 2008, 31716 cases and 141 deaths were reported. A study in a predefined endemic focus revealed that an optimal of 8 Man-Hour-Density (MHD) which is critical vector density was essential for Ph. argentipes enabling it to transmit infection from one infected host to a new host. The transmission cycle of Kala-azar vector is shown in fig. 1. The sand fly is generally confined in the dwellings and occasionally goes out to take sucrose from the soft stem plants in the vicinity to digest the blood meal. In the present study, it was observed...
that the preponderance of vector (except winter months) was 2 to 3 times higher than the critical density in summer and rainy months (from March – October) thereby suggesting probability of successful transmission subject to presence of host and conducive environmental situation,[2, 4, 6, 8, 11]

1.2 Study area and Materials

In the previous study, endemic area of Vaishali district (5 villages) and non-endemic area of Lohardaga district (5 villages) have been taken and studied in detail for understanding geo-environmental parameters. Based on the encouraging results, the endemic area of Vaishali district covering 70.08 sq. km comprising 12 blocks of all villages have been taken for the present investigation (Fig. 2). These villages mostly have mixed dwelling (house with cattle shed) with physical characteristics of the houses consisting of mud walls, mud floors, thatched roof etc. (Shreekant et al, 2010). Indian Remote Sensing satellite data of LISS IV (5.8 m) multispectral and Cartosat – 1 panchromatic data (2.5 m) acquired during the year 2008 and 2009 over two seasons (summer and winter) have been used for the present investigation. Village wise Climatic variables viz. temperature (min & max) and humidity have also been collected. Field investigations have been carried out and information on disease incidence, Man-hour-density (MHD) i.e vector density measure, peri-domestic vegetation, housing pattern, nature of dwellings, and living conditions of the population have been collected.

![Fig. 2 Study area](image)

![Fig. 3 Housing pattern near Gurhi village showing peri-domestic vegetation and mixed dwellings.](image)

During the field investigation, it was observed that most of the villages are having human dwellings with thatched roofs, mud walls with peri-domestic vegetation (Fig. 3).

1.3 Methodology

A systematic approach has been designed based on the information generated using satellite data, field and collateral data and a software package has been developed to estimate villages at risk in the endemic area. The overall methodology is given in Fig. 4. Land use / Land cover, Normalised Vegetation Index, wetlands, rural built up/settlement, peri-domestic vegetation etc. have been derived from satellite data using digital image processing techniques with reasonable accuracy. The understanding of sand fly life cycle became an important input for characterising geo-environmental parameters in micro scale in and around endemic villages. The overall methodology is divided into four different steps namely satellite data processing, collateral data analysis, statistical analysis and GIS model development for early warning system.

**Satellite data processing:** IRS LISS IV and CARTOSAT -1 satellite data pertaining to two seasons have been rectified using reference maps and the study area has been extracted. Different digital enhancement techniques like linear enhancement, histogram equalisation etc. have been applied on the satellite data. IRS LISS IV and CARTOSAT -1 have been merged using Brovey transform to increase interpretability. Land use / land cover has been generated applying supervised classification technique using two seasons satellite data. False colour composite of IRS LISS IV shows numerous wetlands in the study area (fig.5). Normalised Difference Vegetation index has been performed to assess the vegetation vigour of the area in two seasons.

**Collateral data analysis:** Temperature (Max & Min) and humidity data, Kala azar disease incidence, cropping pattern, chemical analysis of soil data and Man-hour-density of vector (MHD) from most of the villages of Vaishali district have been collected and analysed. The analysis of MHD and disease dynamics indicate periodicity at seasonal and inter-annual temporal scale. From 2003 to 2007, there has been a steady
increase in the annual number of disease cases. The Annual vector density of *Aedes aegypti* is reflected in the form of two peaks. Disease epidemic starts at summer and peaks in the March - April and major peak is seen in August-September [1,3]. Dry and humid with higher temperature is favourable climatic condition to the vector in the endemic area. Table - 1 shows population affected and death occurred in twelve blocks of Vaishali district which is reported in Public Health Centres during the years 2008-2010[9]. The population at risk in Mahua block indicates the maximum of 419, 198 and 191 respectively during the years 2008-2010. It is also observed that a large number of disease incidence is not reported in the PHCs regarding disease incidence of the study area.

**Table 1- Disease incidence in various Blocks / PHC of Vaishali District**

Multivariate Analysis: A Systematic study of geo-environmental parameters derived from satellite data in conjunction with ground intelligence enabled modelling of infectious disease and risk villages. High resolution Indian satellites data of IRS LISS IV (multi-spectral) and Cartosat-1 (Pan) have been used for studying environmentally risk parameters viz. peri-domestic vegetation, dwelling condition, wetland ecosystem, cropping pattern, Normalised Difference Vegetation Index (NDVI), detailed land use etc towards risk assessment. Univariate analysis of the relationship between vector density and various land cover categories and climatic variables suggested that all the variables are significantly correlated. Using the significantly correlated variables with vector density, a seasonal multivariate regression model has been carried out incorporating geo-environmental parameters, climatic variables and seasonal time series disease parameters. Linear and non-linear models have been applied for periodicity

Multivariate regression analysis has been carried out using geo-environmental parameters and climatic variables and derived coefficients are shown in equation below.

**Multivariate Regression Equation for Vector MHD**

\[ Z = -42.23 + (\text{Temp.} \times 0.597) + (\text{Humid.} \times 0.684) - (\text{Dry fallow} \times 0.170) + (\text{Min. NDVI} \times 11.44) \]

Where, \( Z \) is the estimated man-hour-density.

The statistical analysis of the parameters indicate that average maximum temperature, humidity, settlement with vegetation, dry fallow, moist fallow, minimum NDVI are significantly correlated with MHD (Table-2).

**GIS Model:** The patterns in the dynamics of the vector diseases are their periodicity and seasonal and inter-annual temporal scale forms the basis for the development of Early warning system [7]. Satellite data forms the primary input for generation of geo-environmental parameters. The following data sets have been generated using satellite data processing /digitisation and organised in the GIS environment(Table 3).

**Table 3 – List of variables used for the analysis**

To improve the MHD predictive approach, fuzzy model has been developed and incorporated in GIS environment combining spatial geo-environmental and climatic variables by applying a Boolean logic for the members of the membership function and the equation is given below.

The combined Fuzzy membership function is defined as

\[ \mu = (\text{Fuzzy algebraic Sum})^\gamma * (\text{Fuzzy algebraic Product})^{1 - \gamma} \]

Fuzzy algebraic sum = \[ \prod_{i=1}^{n} (1 - \mu_i) \]

Fuzzy algebraic product = \[ \prod_{i=1}^{n} (\mu_i) \]

where \( \mu_i \) is the fuzzy membership for \( i^{th} \) map and \( i = 1,2,3,\ldots \) maps are to be combined and where \( \mu_{\text{combined}} \) fuzzy membership of \( Y \) operation, \( \gamma \) is the parameter chosen in the range of (0,1) such that when \( \gamma = 1 \), the output of combined fuzzy membership value equals to same as fuzzy algebraic sum and when \( \gamma = 0 \), the combination equals to the fuzzy algebraic product. Therefore, a judicial choice of \( \gamma \) factor produces output values that ensure the compromise between increasing tendencies of the fuzzy algebraic sum and the decrease effect of the fuzzy algebraic product. Jeyaram, (2008) has used the above techniques successfully in prioritising water conservation
Predicted MHD correlated variables with MHD, stepwise multivariate linear regression analysis was carried out to establish the relationship between predictor variables affecting vector density. The variables (temperature, humidity, dry fallow, min. NDVI) were found to be the best predictor of vector density, as analyzed by linear regression. The regression model indicate that measured and predicted MHD shows close agreement and $R^2$ value calculated as 0.78. This indicates that model has predicted MHD reasonably well. In Gurhi, Inta and Kujji villages have indicated higher MHD based on the geo-environmental parameters (Fig. 6).

A software package for modelling the risk villages integrating multivariate regression and fuzzy membership analysis models have been developed using .NET environment and OPEN Source GIS and Image Processing tools (Dotspatial, Freeimage) to estimate MHD (vector density) as part of the early warning system. The software package has an integrated GIS and Image Viewer with specific Kala-Azar Vector Density prediction and Fuzzy based Kala-Azar Risk assessment modules. These models have been applied successfully through the software package in highly endemic Vaishali district, India and vector density have been calculated with good accuracy and correlated with Kala-azar disease incidence in the district. Risk modelling of villages and Early Warning System developed in coordination with Rajendra Memorial research Institute of Medical Sciences, Patna, India provided predictive measures of MHD-vector density in different villages and different seasons with reasonably good accuracy and maximize the surveillance and control strategy.

1.4 Results and Discussions

Kala-azar risk villages in Vaishali district of Bihar have been predicted based on multivariate regression models and fuzzy based GIS model. It is observed higher degree of vector density have been estimated in the villages where wetlands are associated with good vegetation and higher humidity. MHD is more in villages of mixed dwellings with mud walls and thatched roofs. It is also observed that monthly mean maximum temperature below 37.8°C and monthly mean minimum temperature above 7.2°C, a mean annual relative humidity of 70% or more with a level not falling below 80% for at least three months, an annual rainfall of 1250 mm or more with favourable altitude below 600 m, soft stem peri-domestic vegetation (banana), alluvial soil, high subsoil water level and abundant vegetation as the most favourable factors [10]. All the above ecological conditions prevail in most part of the northern and central region of Bihar and highly suitable for abundance of sandflies in these regions facilitating perennial transmission of kala-azar.

_P. argentipes_ prefers to rest in indoors, about 8-10 times higher in cattle dwellings than in human dwellings. The most favoured resting sites for sandflies include soil cracks, crevices inside the human dwellings and cattlesheds. The species is predominantly endophagic as evident from their higher indoor collections.

Temperature, Humidity, man-hour-density of sandfly, peridomestic vegetation was collected from field based observations. Soil samples collected from fields were also analyzed for its constituents. Landuse parameters and NDVI values were derived from satellite images. The results of the ground - sampling survey and the village wise statistical point data extracted from the landuse and NDVI data were analyzed using an ingeniously developed software module developed for linear multivariate regression analysis. Using significantly correlated variables with MHD, stepwise multivariate linear

![Observed vs Predicted MHD](image)

### Fig. 6 Graph showing comparison of observed and predicted MHD

![Village locations](image)

### Fig. 7 Village locations (a) Observed MHD and (b) predicted MHD overlaid on IRS satellite data of Patepur Block, Vaishali District.

The calculated MHD using the regression model and fuzzy based GIS model is overlaid on the satellite data of Patepur Block, Vaishali District which indicates the prevalence of higher MHD in village locations where more wetlands and good vegetation exists (Fig. 7).

The Standalone Software package (developed using Open source GIS and Image Processing tools) is successfully customised for Kala-Azar Risk modelling which has been used for deriving the predicted MHD. For each point of interest (in this case settlement locations) Landuse, NDVI and other climate predictors are extracted in its proximity to determine the suitable conditions required for the high vector density. Thus the predicted MHD is calculated using multivariate regression equation. The Software package has easy to use interface, supporting various data formats (GIS, Image, Non –spatial data format) and custom modules like Data creation / Editing.
Reporting, Charting/statistics, Image Geotagging and Geoprocessing (Fig. 8).

![Fig. 8 Main Viewer of Software package](image)

1.5 Conclusion

The multivariate regression analysis between MHD and the predictor variables was found to hold true for the given epidemic area. Thus this statistical relation can be used in these regions for disease forecasting. This model has been applied successfully through software package in highly endemic Vaishali district, India and vector density have been calculated with good accuracy and correlated with Kala-azar disease incidence in the district. Risk modelling of villages and Early Warning System developed in coordination with Rajendra Memorial research Institute of Medical Sciences, Patna, India provided predictive measures of MHD-vector density in different villages and different seasons with reasonably good accuracy and maximize the surveillance and control strategy.

1.6 References


