Geospatial Health 9(1), 2014, pp. 179-191

Environmental modelling of visceral leishmaniasis by susceptibility-mapping using neural networks: a case study in north-western Iran

Mohammadreza Rajabi[1], Ali Mansourian[1], Petter Pilesjö[1], Ahad Bazmani[2]

[1]Lund University GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden; [2]Infectious and Tropical Diseases Research Centre, Tabriz University of Medical Sciences, Tabriz, Iran

Abstract. Visceral leishmaniasis (VL) is a potentially fatal vector-borne zoonotic disease, which has become an increasing public health problem in the north-western part of Iran. This work presents an environmental health modelling approach to map the potential of VL outbreaks in this part of the country. Radial basis functional link networks is used as a data-driven method for predictive mapping of VL in the study area. The high susceptibility areas for VL outbreaks account for 36.3% of the study area and occur mainly in the north (which may affect the neighbouring countries) and South (which is a warning for other provinces in Iran). These parts of the study area have many nomadic, riverside villages. The overall accuracy of the resultant map was 92% in endemic villages. Such susceptibility maps can be used as reconnaissance guides for planning of effective control strategies and identification of possible new VL endemic areas.

Keywords: visceral leishmaniasis, environment, geographical information systems, neural networks, Iran.

Introduction

Visceral leishmaniasis (VL), also known as kala-azar, is a zoonotic, vector-borne disease, endemic in 76 countries. This disease is the second-largest parasitic killer in the world (after malaria), responsible for an estimated 500,000 infections each year (Desjeux, 2001; Palit et al., 2005). Between 20,000 and 40,000 of people die from VL annually (Alvar et al., 2012). According to World Health Organization (WHO) reports, VL is a neglected tropical disease (http://www.who.int/neglected_diseases/2012report/en) that might spread throughout developing countries, where the conditions for the disease exist. Without any control strategies, the untreated VL can have a fatality rate as high as 100% within 2 years (http://www.who.int/leishmaniasis).

Leishmania infantum is the principal agent of human and canine VL in Iran (Mohebali et al., 2002, 2004). The disease has been reported sporadically in Iran with the north-western and southern part being the primary, endemic foci, where it is most frequently seen among the rural population and nomads. From 1996 to 2010, more than 3,000 cases of symptomatic VL were detected in 31 of Iran's provinces. The majority of the cases (92.8%) were found among children up to 12 years old (Mohebali, 2012).

The social and physical environment of north-western Iran is characterised mainly by the presence of several factors strongly associated with VL, including nomadic lifestyle, suitable climatic conditions and large dog populations (i.e. sheepdogs, guard dogs and stray dogs) (Edrissian et al., 1988; Mirsamadi et al., 2003; Moshfe et al., 2008; Salahi-Moghaddam et al., 2010; Rajabi et al., 2012). The Meshkin-Shahr district in the Ardabil province is one of the most important endemic zones for VL in north-western Iran (Salahi-Moghaddam et al., 2010) and the Iranian Ministry of Health (MoH) records indicate that the number of VL infections has noticeably increased in this and neighbouring districts during the last decade (Soleimanzadeh et al., 1993; Tamook et al., 2006). Newer reports shows that VL occurs also in other provinces in north-western Iran, including East Azerbaijan (Mirsamadi et al., 2003; MoH, 2006, 2008). Further investigations have shown that VL has become common in two districts of East Azerbaijan, i.e. Kalaybar and Ahar, which have the largest VL-infected population in the province (MoH, 2006, 2008; Fallah and Farshchian, 2009; Khanmohammadi et al., 2010). The study focused on 800 villages in these two districts: Kalaybar in the north-eastern part of East Azerbaijan and Ahar, located immediately south of Kalaybar (Fig. 1). The above-mentioned

Corresponding author:
Ali Mansourian
Lund University GIS Centre
Department of Physical Geography and Ecosystem Science
Lund University, Lund, Sweden
Tel. +46 46 222-1733; Fax +46 46 222-0321
E-mail: ali.mansourian@nateko.lu.se

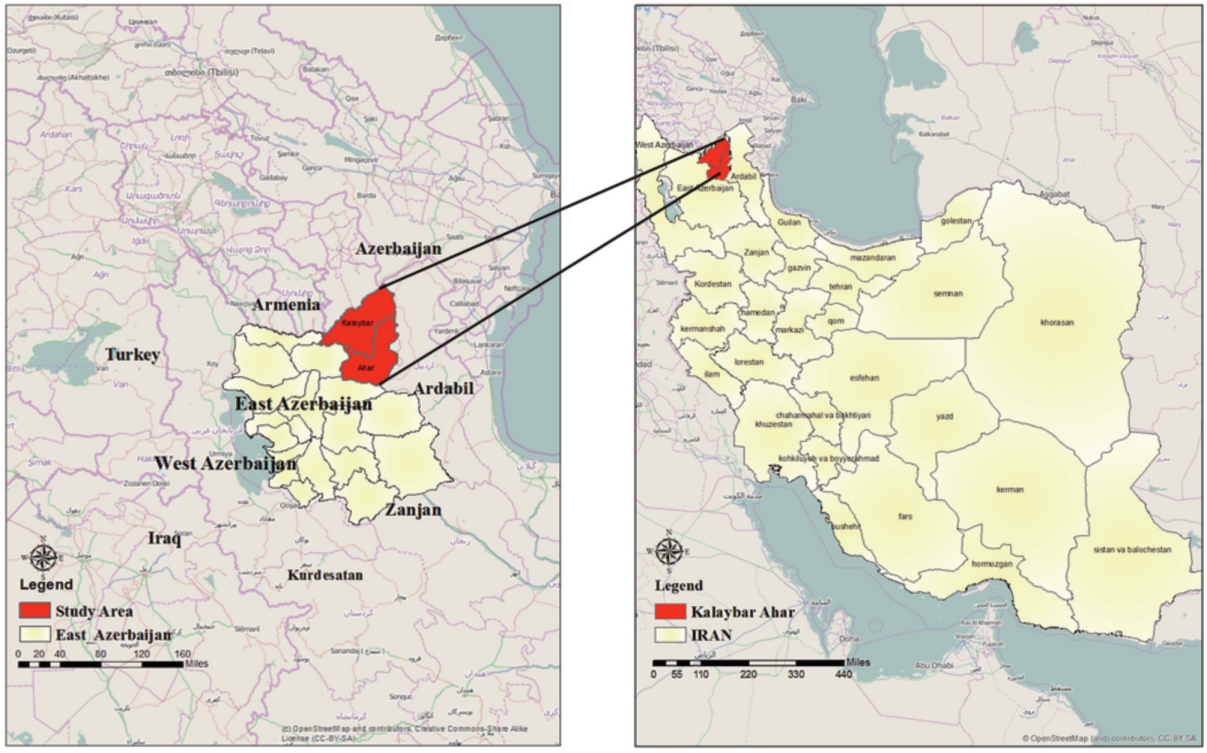


Fig. 1. Study area in East Azerbaijan, Iran.

Iranian endemic areas border three other countries: Azerbaijan, Armenia and Turkey, raising substantial national and international concern over the probability of international spread.

A number of studies around the world have demonstrated that environmental, demographic and statistical data about the ecology of VL can provide the basis for the development of spatial, predictive risk models (Peterson and Shaw, 2003; Castillo-Riquelme et al., 2008; Salahi-Moghaddam et al., 2010). Geographical information systems (GIS) can be applied for risk mapping and identifying endemic areas for diseases (Elnaiem et al., 2003; Rinaldi et al., 2006; Pezeshki et al., 2012). For example, Bhunia et al. (2011) studied the influence of the distribution of inland water bodies on transmission of VL and presence of its dominant vector in a GIS model. In a knowledge-driven approach, Rajabi et al. (2012) used multi-criteria decision-making methods together with GIS to identify high-risk areas for VL outbreak in north-western Iran. The techniques applied can be divided into two groups: data-driven and knowledge-driven ones. In the former approach, the analyst uses expert knowledge to assign weights to a series of factors. The main problem here is that insufficient knowledge could mislead the work. Another limitation is the subjectivity of the weighting of the factors. Deterministic, data-driven methods can only be applied in areas where there are simple relations between the predetermined factors and the desired phenomena. Moreover, the environmental properties should not be inhomogeneous and intricate (Turner and Schuster, 1996). In addition, these methods are only applicable for small areas (Yilmaz, 2009). One of the main drawbacks of the deterministic, data-driven methods is malfunctioning when the data are incomplete (Gomez and Kavzoglu, 2005). Statistical, data-driven methods, on the other hand, require the collection of large amounts of data to produce reliable results (Yilmaz, 2009).

To overcome these limitations, new techniques, such as artificial neural networks (ANNs) are used for modelling complex problems. These advanced, data-driven approaches are being used more frequently in spatial studies to identify and classify areas as well as to predict the distribution of real-world phenomena having to do with flooding, landslides, mines, among others (Biswajeet and Saro, 2007; Nykanen, 2008; Yilmaz, 2009; Choi et al., 2010; Pradhan, 2013). The ANN algorithm imitates the structure of the human brain and conducts predictions based on repeated trainings (Yeo and Yee, 2014). The ANN architecture consists of a number of interconnected neural neurons that are modelled by mathematical functions (Haykin, 1998). ANN learns from experience and can be trained to

recognise patterns, classify data and forecast future events (Kohonen, 1996; Ripley, 1996; Bishop, 1997; Matouq et al., 2013). This makes ANN a powerful tool for modelling purposes, especially when the underlying data relationships are unknown (Lek and Guegan, 1999; Matouq et al., 2013).

Although VL is restricted to specific localities with special environmental, topographical, demographical and socioeconomic factors, there has been no study on adapting ANNs in GIS to explain the focal distribution of VL. ANNs have indeed been relatively poorly utilised in spatial epidemiological studies. Kiang et al. (2006) used ANN methods to model the dependency of malaria transmission on precipitation, temperature, relative humidity and vegetation index variables in Thailand. Capinha et al. (2009) also applied ANN to combine *Anopheles atroparvus* species records with a set of five environmental predictors to present habitat suitability for malaria vectors in Portugal. Moustris et al. (2012) provided an ANN forecasting model to evaluate the possible impact of meteorological parameters and air pollution on the number of childhood asthma admissions. To our knowledge, no such studies have applied ANNs to model VL prevalence.

The main aim of this study was to develop a model based on ANN to map the potential risk-prone areas in relation to VL outbreak in the study area. Since VL is a vector-borne disease, which spreads mostly by means of reservoir hosts (e.g. dogs), it is not possible to collect complete input data for the model. ANN represents a powerful data-driven approach that models the behaviour of VL vectors and reservoir hosts based on a sample data collection investigating the complex association between the environmental properties of the study area, the socioeconomic factors and the spread of VL. With this in mind, we aimed to develop an environmental model for predictive mapping of the most susceptible areas for VL, taking into account meteorological, topographical, demographic and socioeconomic factors. The map would provide new insight, which would help to develop strategies for preventing further spread of VL in the study area. To specify the input data for the prediction model, the association between VL endemicity and various factors in the literature was explored. In this regard, seven items were deduced from literature to be the critical recognition criteria (CRC) for predictive mapping of VL risk-prone areas: temperature, precipitation, proximity to rivers, altitude, presence of health-centres, land cover and presence of nomads (Sudhakar et al., 2006; Bhunia et al., 2010; Salahi-Moghaddam et al., 2010; Rajabi et al., 2012; Tsegaw et al., 2013).

Materials and methods

Radial basis functional link nets

ANNs mimic the nervous systems of the human brain. They can be defined as simplified mathematical models trained to learn (Tsoukalas and Uhrig, 1997; Beucher et al., 2013). The fundamental elements are neurons, which receive multiple signals, combine and modify them to transmit the result to other neurons. In an ANN, the artificial neurons are usually organised in layers (Beucher et al., 2013). In this study, we used a neural network model called radial basis functional link nets (RBFLN) described by Looney (1997, 2002). This specific model was chosen since, unlike other ANNs, it requires a smaller volume of training data (Looney, 1997, 2002). In spatial epidemiology studies, especially in poorly explored areas, the number of known areas susceptible to a specific endemicity like VL is low (i.e. the amount of training data that can be obtained is low). One of the advantages of RBFLN, desirable for disease predictive mapping, is that a smaller number of hidden nodes can be used and hence good results can still be achieved even with a small amount of training data (Looney, 1997, 2000; Looney and Yu, 2001).

An RBFLN is composed of three layers (Fig. 2): (i) an input layer of N nodes, where each node has one input (e.g. an area in a map layer) representing a feature vector of N elements; (ii) a hidden layer of M artificial neurons, where each neuron signifies a radial basis function (RBF); and (iii) an output layer of J artificial neurons (Looney, 1997; Looney and Yu, 2001, Beucher et al., 2013). In GIS-based applications the feature vectors are obtained by the combination of N evidential map layers. Each feature vector is fed to the input layer and transmitted to the hidden layer. The sum of inputs, x (x_1, x_2, …, x_N), in each neuron in the hidden layer is then fed into a radial basis function, which results into a single output y. The output values of the hidden layer, y (y_1, y_2, …, y_M), should be multiplied by weights u_{mj} while being transmitted to the output layer. Moreover, the output layers in an RBFLN also receive the inputs x multiplied by a second set of weights w_{nj}, directly. Finally, each output layer neuron returns a unique output z (corresponds to a specific area with a specific endemicity) which is generated using the values of inputs x, input weights w_{nj}, the outputs of hidden layer y, and their weights u_{mj}.

Every ANN method has a self-organizing structure which enables the system to learn from experience (i.e. training data, which generates target vectors). During

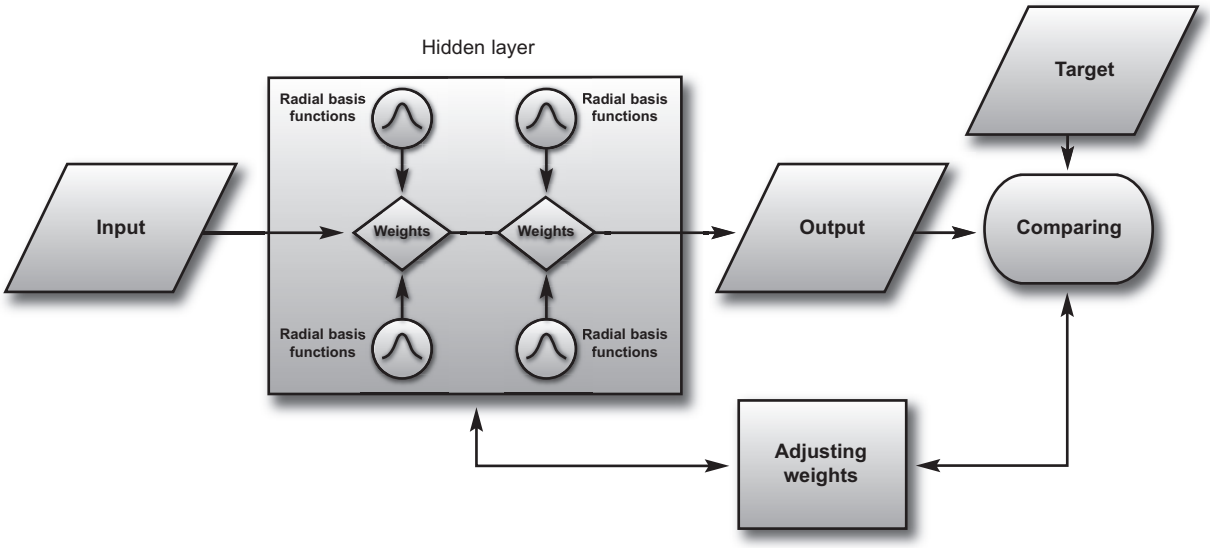


Fig. 2. The general process of the RBFLN algorithm as an ANN.

the training process a probability value (z) is calculated for each target vector (t) by progressive iterations between neurons. The weights, u_{mj} and w_{nj}, control the training by affecting the interneuron connections. Hence, the weights are repeatedly modified until the outputs z (z_1, z_2, …, y_J) approach the targets t (t_1, t_2, … , t_J). The RBFLN is capable of accounting for model linear and nonlinear input-output relationships (Looney, 2002). In this regard, since there are a variety of intricate relations between the disease endemicity and the environment, we find this method very useful for spatial epidemiology. Another advantage of RBFLNs is that a smaller number of hidden neurones can be used, which in turn increases the speed (fewer iterations) and accuracy of the learning (Looney, 2002).

Data resources

The data input were gathered/generated from different sources (Table 1). Topographical and environmental maps at the scale of 1:50,000 were used to obtain altitude and river data together with village information. Data on rainfall and temperature were extracted from mapping of the data gathered by the Iranian Meteorological Organization. Seven multi-class map layers were produced (Fig. 3) based on available village-level data including demographic, socio-economic, environmental and healthcare access data were linked to the relevant villages using GIS. VL notification data were acquired from the Infectious and Tropical Diseases Research Center (ITDRC) of the MoH.

Table 1. Summary of data used for VL modelling in RBFLN.

Data	Evidence map	Source
Topographical	Altitude Proximity to rivers	NCC*, 1:50,000 maps
Meteorological	Precipitation Temperature	Meteorological organization
Environmental	Nomadic villages Proximity to health centers	NCC*,1:50,000 maps
Land cover	Land cover	Forest and Rangeland Organization
VL notification	Training and validation points	(i) ITDRC of the MoH** (ii) Field data collected by authors with the collaboration of ITDRC

*National Cartography Center; ** Infectious and Tropical Diseases Research Center of Ministry of Health (MoH), Tabriz branch.

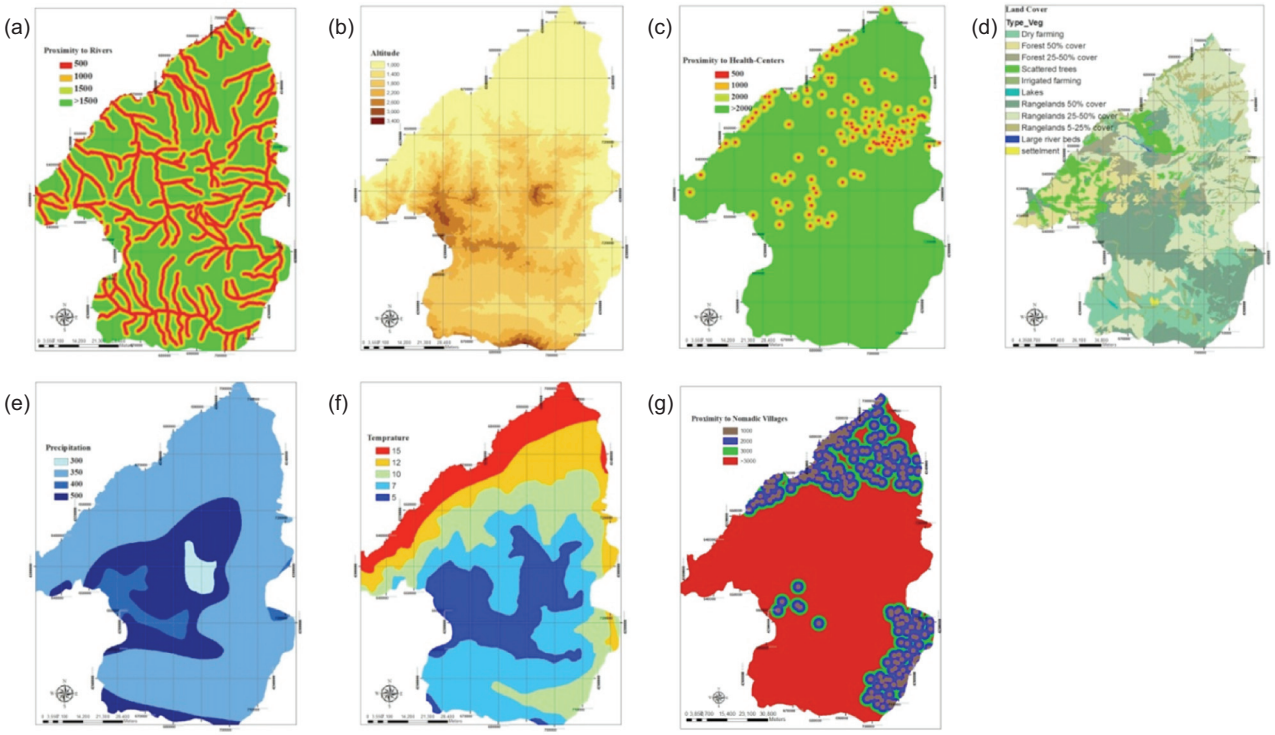


Fig. 3. RBFLN input evidence maps. (a) Proximity to rivers, (b) Altitude, (c) Proximity to health-centres, (d) Land cover, (e) Precipitation, (f) Temperature, and (g) Proximity to nomadic villages.

Training data

The RBFLN network requires two sets of training datasets: one that defines the presence of the objects or conditions to be predicted (i.e. the VL endemic areas) and a second that defines the absence of these objects (i.e. locations where VL is known not to be endemic). The two datasets used to produce training data for presence of VL endemic areas were data on infected villages and data on VL reservoir hosts through study area.

The former dataset was used as the primary reference to identify infected areas (Fig. 4a). Detailed characteristics of VL incidence were investigated and then 17 highly infected villages were entered. Second, a local survey was performed to acquire more insight into VL canine infections. In this regard, a dataset about the degree of VL infection in dogs through the study area was collected based on two distinct surveys. In the first survey, sheepdog and guard dog population data were acquired by nomad census (Fig. 4b) using three different data-collector teams. In the second survey, data on the dispersal of stray canines were gathered by irregular data gathering of five data-collector teams. As a result of these two surveys, data for 400 canines from 19 villages were gathered. To identify VL infections in the canines, direct agglutination test (DAT) was performed in collaboration with ITDRC; 130 out of 400 canine tests were positive (Fig. 4c). Information from 13 villages, where the DAT results indicated a prevalence of >10% canine infections, were entered into the training sets as endemic areas (Fig. 4d). Therefore, 30 (17+13) endemic points were generated as training data for VL presence. Five points out of 30 from different parts of the study area were selected as endemic validation points and not used as input for the RBFLN model.

To acquire the second subset of RBFLN training data, in which the absence of VL endemicity should be determined, the areas where the VL outbreak has a low risk were chosen using MoH records. Investigations showed that there had been no VL incidence over the last 20 years in 57 of the villages in the study area. Since the flight reach of the VL vector, the sand fly, does not exceed 1 km, this measure was used as the base distance when exploring the neighbourhood of the villages. Accordingly, spatial analysis indicated that there were no infected villages within the 1 km radius of 48 out of the 57 villages. Hence, neighbourhoods up to 1 km away from these 48 villages were considered as VL-safe zones. We chose 26 random points in these areas with five of them used as non-endemic validation points. These non-endemic validation points were not used as input data for the RBFLN model.

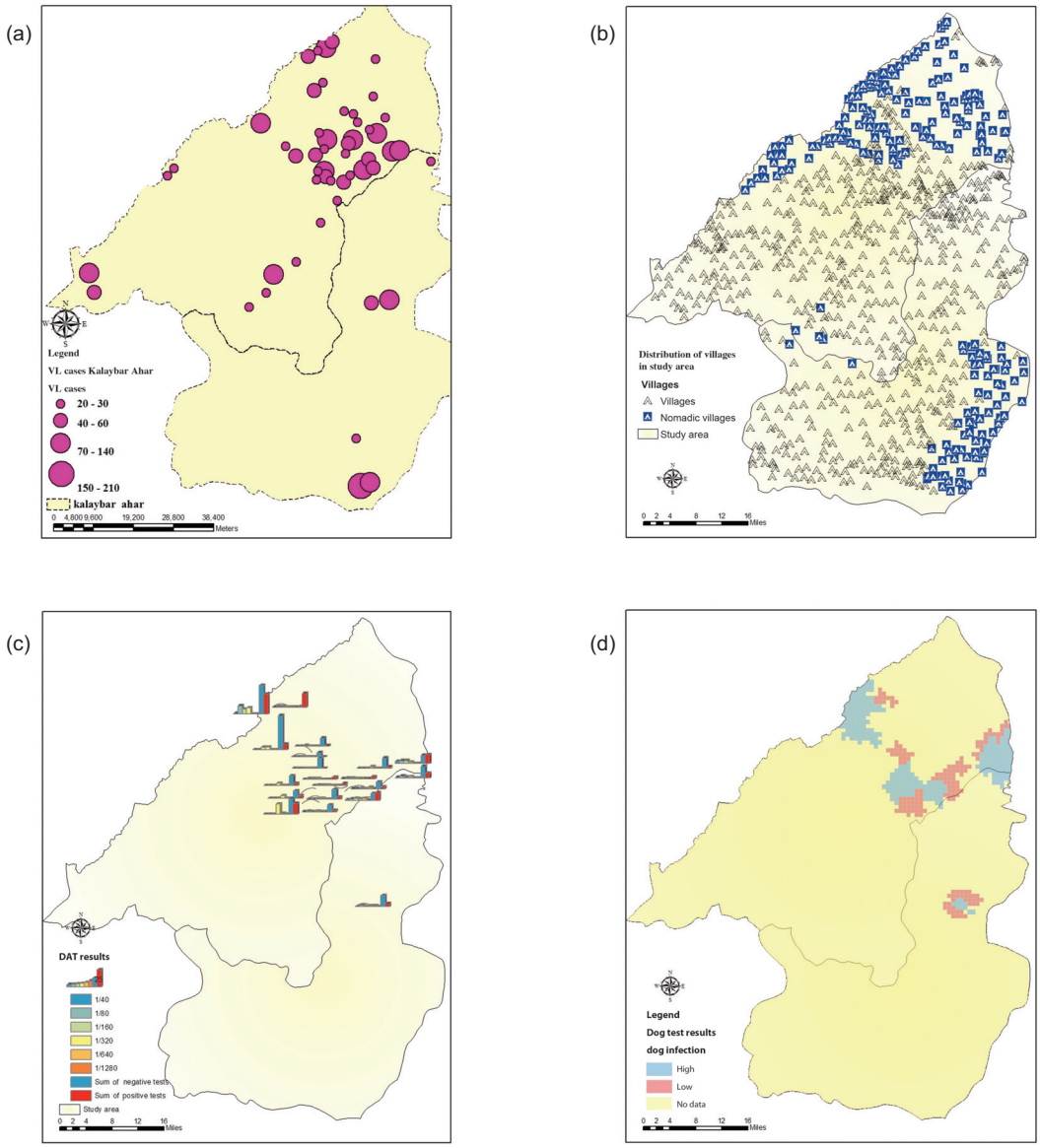


Fig. 4. Primary data used for VL endemic training data: (a) registered VL infection cases; (b) nomadic villages; (c) DAT test results; (d) canine infection by VL in explored areas.

RBFLN model for VL

Data pre-processing and integration were performed using ArcGIS version 10 (ESRI; Redlands, USA) and an extension for multivariate analysis called spatial data modeller (SDM) (Sawatzky et al., 2009). The input data used in the neural network model, including the multi-class, evidential maps of rainfall, temperature, altitude, proximity to rivers, proximity to health centres, proximity to nomadic villages and land cover, were received as input data as feature vectors by the RBFLN. To generate a feature vector map for RBFLN in the model, the maps were overlaid to create a "unique conditions grid map". This map is defined as an integer grid formed by the combination of two or more predictor maps, in which the class values represent uniquely-occurring combinations of the classes of the input themes according to Kemp et al. (1999). A "unique conditions grid" consisting of 2,699 feature vectors was generated. In the attribute table of this grid, there is one row for each unique overlay condition as well as one column for each evidential map. In this sense, each feature vector referred to a specific geographical situation through the study area. Thus the unique overlay conditions are N-dimensional (N = number of evidential maps) vectors.

The training data set in RBFLN contains Q feature vectors and Q associated output target vectors $\{x_q : q$

= 1, 2, ..., Q} and {t_q : q = 1, 2, ..., Q}. For the VL RBFLN model, the feature vectors spatially-coincident with the locations of the 46 training and 10 validation endemic and non-endemic areas, were extracted to generate training and validation samples, respectively. In this research, there were 46 training points, Q = 46, and all of the map values were normalised in a [0,1] range using $x_n = (x - x_{min}) / (x_{max} - x_{min})$ where x_n is the normalised value of x, x_{max} the maximum value of the variable x and x_{min} the minimum value of x. For example, the first feature vector in our training was: $x^{(1)}$: {0.0, 1.0, 0.33, 0.4, 0.66, 0.33, 1.0} and $t^{(1)}$ {1.0}. Each element in the $x^{(q)}$ vector corresponds to a normalised value of a specific area in each of the seven predicator maps. The element of $t^{(q)}$ determines whether this vector shows the presence (1.0) or the absence (0.0) of the desired event (i.e. VL-infected areas).

The first stage of training in the RBFLN model is to initialise the values of the parameters in the mathematical functions. These parameters include centres (v), spread parameters (σ) and synaptic weights (u, w). In this research the initial value of σ was calculated using the equation:

$$\sigma = (1/4)(1/M)1/N \quad \text{(equation 1)}$$

where M is the number of hidden neurons (radial basis functions) and N the number of the predicator maps. At the next stage, the assigned synaptic weights and spread parameters would be adjusted to minimise the difference between the model output (z) and target values (t) meaning that the variables u and w were changed during the training process until the predicted VL risk values of known endemic and non-endemic areas reaches their predetermined risk values, i.e. the model is learning where to find VL-susceptible areas. The better the model learns, the more accurate the values of synaptic weights (u, w) become. The output values of RBFLN are calculated by as follows:

$$z_j^q = [1/(M + N)] \left[\sum_{m=1}^{M} u_{mj} \times y_m^q + \sum_{n=1}^{N} w_{yj} \times x_n^q + b_j \right] \quad \text{(equation 2)}$$

where

$$y_m^q = e^{[-\|x^q - v^m\|^2/2\sigma_m^2]}$$

For the purpose of VL modelling using RBFLN, an optimum structure for RBFLN in terms of the number of hidden neurons as well as the number of iterations for RBFLN training must first be determined. To do so, RBFLNs with different structures were explored. A training total sum squared error (SSE) was recorded for each structure/run: the lower the SSE, the better RBFLN structure (Fig. 5). An RBFLN structure with 45 hidden functions (M = 45) and 1,000 iterations, resulting in a summed-squared error (SSE) equal to 0.0038 was considered to be the best structure among those considered (Fig. 6).

Validation of RBFLN model for VL

The validation of RBFLN results for VL predictive mapping was performed based on two different approaches. First, a receiver operating characteristic (ROC) curve was used to assess the performance of RBFLN model for VL. A ROC curve is a plot of sensitivity (true positive rate: TP/(TP + FN)) on the y-axis *versus* specificity (false positive rate: FP/(FP + TN)) on the x-axis (where TP = true positive, TN = true negative, FP = false positive and FN = false negative). The area under the curve (AUC) can be used as a measure of the classification performance of an ANN (Brown et al., 2003; Obuchowski, 2003). The AUC varies between 0 and 1, the latter signifying a perfectly accurate outcome with 0 indicating that the result is completely incorrect. Second, the VL RBFLN model was validated by plotting the validation points on the curve of RBFLN probability values *versus* cumulative percentages of the study area. Hence, the susceptibility map was validated by plotting the position of endemic and non-endemic validation points on the predictive classification value *versus* the percentage of the cumulative area curve.

Results and discussion

Fig. 7 shows the result of applying the RBFLN to create a multi-class predictive VL map. This map should be interpreted as susceptibility of the individual cells in the area in relation to the VL endemicity. It was used to generate a curve of susceptibility *versus* cumulative percentage of the study area (Fig. 8). Inflection points of the curves allowed the separation of five distinct zones, namely: very low, low, moderate, high and very high susceptibility to VL (Fig. 8). The inflection points were selected where there was a notable stepwise increase in susceptibility in relation to the cumulative area. The cut-offs separating very low from low susceptibility and so on moving towards very high occurred at the susceptibility values of 0.1967, 0.4050, 0.6379 and 0.9775, respectively.

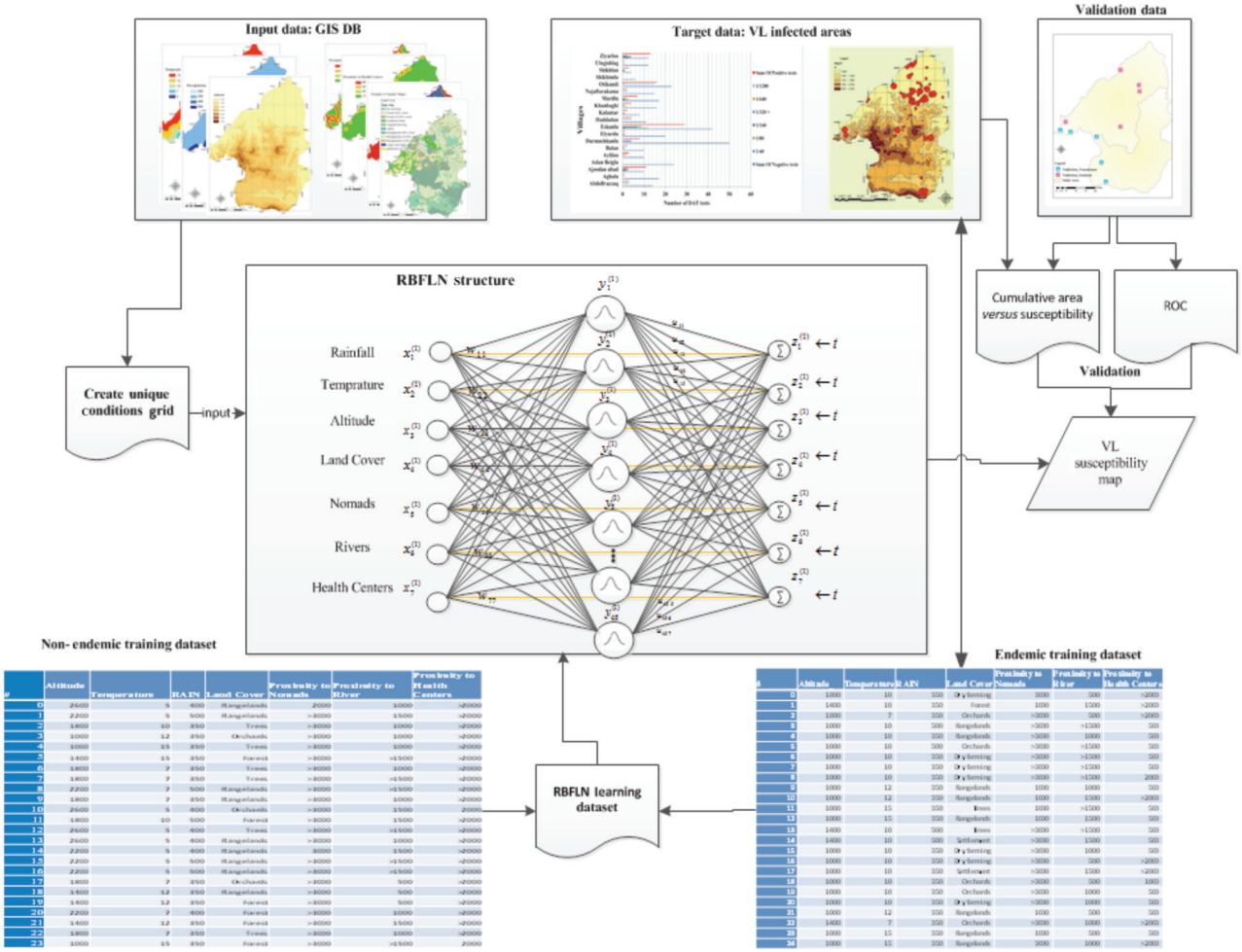
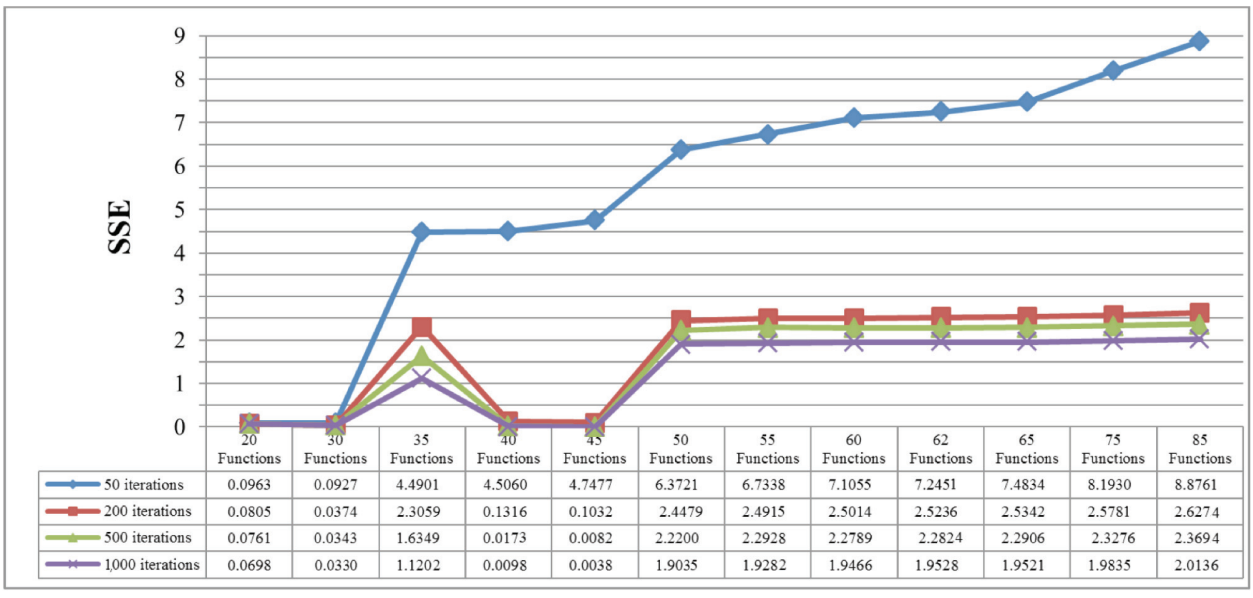


Fig. 5. The general schema of VL-RBFLN model used in this study.



	20 Functions	30 Functions	35 Functions	40 Functions	45 Functions	50 Functions	55 Functions	60 Functions	62 Functions	65 Functions	75 Functions	85 Functions
50 iterations	0.0963	0.0927	4.4901	4.5060	4.7477	6.3721	6.7338	7.1055	7.2451	7.4834	8.1930	8.8761
200 iterations	0.0805	0.0374	2.3059	0.1316	0.1032	2.4479	2.4915	2.5014	2.5236	2.5342	2.5781	2.6274
500 iterations	0.0761	0.0343	1.6349	0.0173	0.0082	2.2200	2.2928	2.2789	2.2824	2.2906	2.3276	2.3694
1000 iterations	0.0698	0.0330	1.1202	0.0098	0.0038	1.9035	1.9282	1.9466	1.9528	1.9521	1.9835	2.0136

Fig. 6. Number of radial basis functions *versus* sum of squared error (SSE) for different training runs indicating minimum SSE for 45 radial basis functions at 1,000 iterations.

Fig. 8 shows the susceptibility values at all training points; 62.5% of training points (representing non-endemic villages) fell into the low and very low susceptibility areas with values <0.405. There were no non-endemic training points in the high susceptibility area. 92% of the endemic villages were placed in the high and very high susceptibility areas. In order to further evaluate the model, the validation data were explored in the map resulting from each training structure. These explorations confirmed that the chosen RBFLN structures had the best performance and choosing any other number of hidden functions did not improve the performance of the neural network (Table 2).

The susceptibility map (Fig. 7) was interpreted as follows: (i) the high and very high susceptibility zones occupied 36.3% of the study area and contained 20% and 100% of the non-endemic and endemic validation points, respectively; (ii) the moderate susceptibility zones occupied 20.2% of the study area and contained 20% and 0% of the non-endemic and validation points, respectively; and (iii) the low and very low susceptibility zones occupied 43.5% of the study area, did not contain any known endemic validation sites and contained 80% of the non-endemic points. These are also shown in Fig. 9. For the further validation of the current study, in terms of the AUC measure of ROC curves, the RBFLN model produces reliable results (AUC = 0.920). The results of the ROC curve validation are shown in Fig. 10, and the AUC values are listed in Table 3.

The correct classification of almost all validation points by the RBFLN demonstrated the ability of the model to recognise risk situations in an input feature vector (i.e. the GIS layers of the study area). The high and very high susceptibility areas of the VL model in the northern and south-eastern part of the study area (Fig. 7) comprise riverside, nomadic villages without health-centres, which have a suitable climate for the VL sand fly vector. The analysis demonstrates that altitude has a relatively strong influence on the prevalence of VL in that high-susceptibility areas for VL were found to be common at low altitudes and most of the highlands had low susceptibility. Investigations on the land cover of the high-susceptibility areas (Figs. 3d and 7) indicate that low density rangelands (i.e. with 5-50% coverage) and dry farming areas comprised most of the high-risk areas. Nomadic lifestyle, which is associated with relatively poor socioeconomic conditions, was also indicated to be a key factor in the spatial distribution of VL risk. Nomadic villages with moderate weather conditions in the study area (i.e. temperatures between 10 and15 °C, and average precipitation 350 mm) were found to be more prone to VL infection. Additionally, the humidity caused by rivers (Figs. 3a and 7) could be considered a catalyst for VL outbreaks in moderate weather conditions. Meanwhile, accessibility to water is an influential factor for farmers and ranchers, and hence, population density is concentrated near rivers; high population density could be another factor increasing the probability of VL endemicity.

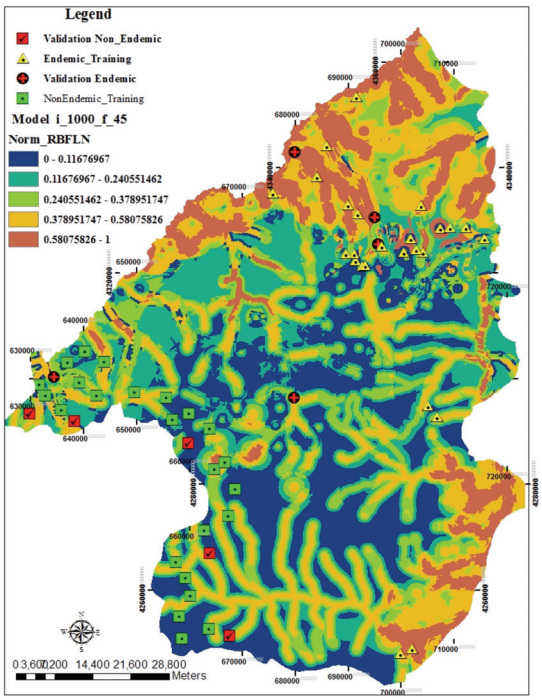


Fig. 7. RBFLN multiclass predictive map for VL.

In order to determine an overall measure of the correlation between the predictive results created by our neural network analysis and those created by other

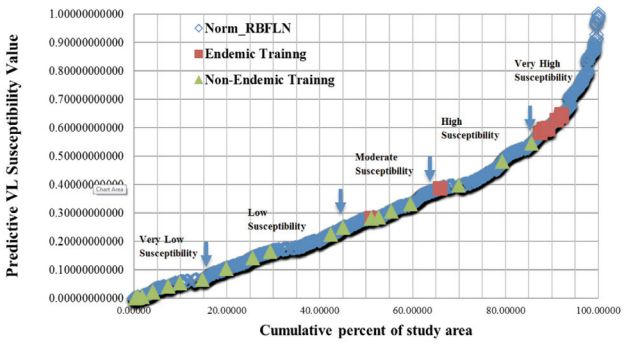


Fig. 8. Susceptibility *versus* cumulative percentage of the study area for the VL model. Arrow lines crossing inflexion points were used to delineate five specific classes of very low susceptibility, low susceptibility, moderate susceptibility, high susceptibility and very high susceptibility.

Table 2. Summary of training for various radial RBFLN structures.

Model name	Parameters				Training		Validation % of correctly classified validation samples	
	UC	x	u	i	MSE	SSE	Endemic	Non-endemic
I_1000_f_20	2,699	46	20	1,000	0.0698	3.2120	20	20
I_1000_f_30	2,699	46	30	1,000	0.0330	1.5190	40	60
I_1000_f_35	2,699	46	35	1,000	0.0244	1.1202	60	80
I_1000_f_40	2,699	46	40	1,000	0.0002	0.0098	80	100
I_1000_f_45	**2,699**	**46**	**45**	**1,000**	**0.0001**	**0.0038**	**100**	**80**
I_1000_f_50	2,699	46	50	1,000	0.0501	2.2538	20	40
I_1000_f_55	2,699	46	55	1,000	0.0419	1.9282	40	60
I_1000_f_60	2,699	46	60	1,000	0.0423	1.9466	40	40
I_1000_f_62	2,699	46	62	1,000	0.0425	1.9528	40	40
I_1000_f_65	2,699	46	65	1,000	0.0424	1.9521	40	40
I_1000_f_75	2,699	46	75	1,000	0.0431	1.9835	40	40
I_1000_f_85	2699	46	85	1,000	0.0438	2.0136	20	20

UC = number of unique conditions; x = number of input feature vectors; u = number of radial basis functions; i = number of iterations; MSE = mean of squared error; SSE = sum of squared error.

methods, a comparison was made with previous comparable studies, i.e. Correa Antonialli et al. (2007), Bhunia et al. (2010), Fernandez et al. (2010), Salahi-Moghaddam et al. (2010), Sheets et al. (2010), Rajabi et al. (2012) and Fu et al. (2013). In this regard, pairwise comparisons were made between the predictive map resulting from the approach of this study (Fig. 7) and predictive maps resulting from fuzzy modelling methods created by Rajabi et al. (2012), which are almost based on the same map outcomes. The comparison shows that areas with a susceptibility value of >0.75 in maps shown by Rajabi et al. (2012) coincide at least 89% with the high and very high susceptible areas according to the RBFLN approach in the northern part of the study area and 81% in the high and very high susceptible areas in the southern part. Hence, it could be interpreted that the RBFLN truly deduced the VL endemicity in the study area without any external knowledge. The main advantage of the RBFLN model in relation to the knowledge-driven approach used in Rajabi et al. (2012) was the independency of the output results in relation to the selective beliefs of experts. Accordingly, there was only one single output map from the RBFLN model. However, in Rajabi et al. (2012) the degree of risk and trade-off between influencing factors differed based on each

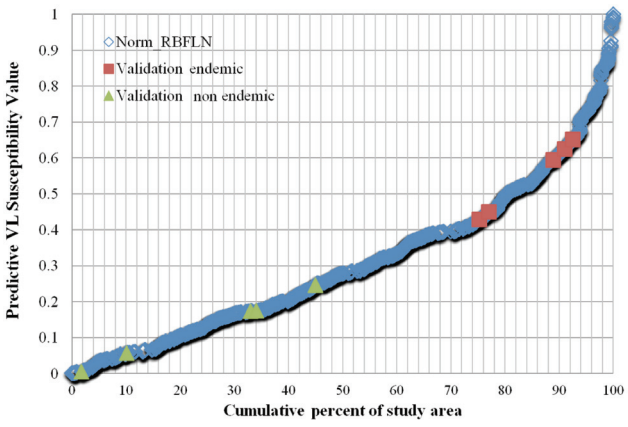


Fig. 9. Variation of predictive VL susceptibility values with cumulative percent of study area. Squares are endemic validation points and triangles are non-endemic validation points.

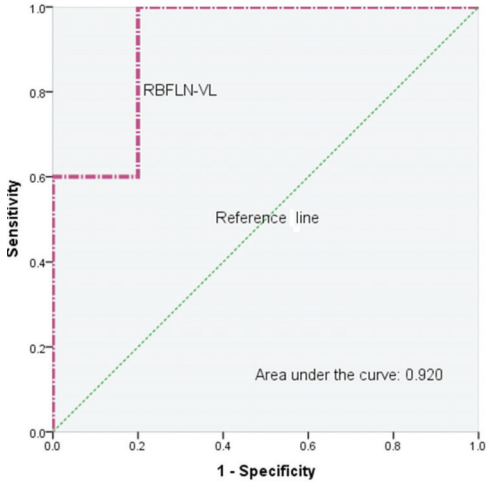


Fig. 10. VROC curve for VL - RBFLN validation.

Table 3. Validation data used for RBFLN.

N.	State	Predicted score	Class
1	Endemic	0.3986	1
2	Endemic	0.4487	1
3	Endemic	0.6510	1
4	Endemic	0.6249	1
5	Endemic	0.5938	1
6	Non-endemic	0.0045	0
7	Non-endemic	0.4549	0
8	Non-endemic	0.0561	0
9	Non-endemic	0.2461	0
10	Non-endemic	0.1726	0

expert's opinion and different map derived from their knowledge. Moreover, finding experts who have a reliable knowledge about all of the determining factors and their relations to endemicity is very difficult. In this sense, the knowledge gap which is another drawback in the knowledge-driven approaches was covered in the RBFLN model.

Similarly, it has been shown by Bhunia et al. (2010) that areas with low altitude, high population density, low vegetation density and low living standard are the most suitable areas for the VL vectors. Sheets et al. (2010) also suggested that areas with a low standard of living and with a large number of rainy days are more prone to VL incidence, while Fernandez et al. (2010) reported that the distribution of VL vectors is associated with land coverage (bushes and trees) and accessibility to health facility services in densely populated neighbourhoods. Correa Antonialli et al. (2007) analysed patterns and surface areas in a spatial approach to infer how VL spreads and concluded that migration of people from endemic to non-endemic regions is the most important factor. Likewise, our investigations emphasised the impact of travel, e.g. nomadic people migrating twice annually. Among other factors, Fu et al. (2013) point out the importance of access to healthcare services and lifestyle, while the correlation of VL with rainfall, humidity, elevation, and temperature found here has been confirmed by Salahi-Moghaddam et al. (2010).

Although it has been demonstrated that several different parameters are associated with VL prevalence, most of the aforementioned studies only examined the impact of a limited set of factors. Besides, most of VL environmental analyses have been performed in non-spatial frameworks and the usage of GIS has been limited to just a simple cartographical presentation of the results. Moreover, the environmental knowledge which can be obtained from observational data has been poorly attended in VL prevalence modelling. In the present study, we have adopted a data-driven approach in which field observations and measurements together with records of MoH have been used for a reliable VL endemicity modelling.

Although the RBFLN model for VL is efficient for susceptibility mapping, it has a limitation inherited from its ANN structure as no satisfactory explanation of its behaviour is offered. To investigate detailed behaviour of the model, one can use other artificial intelligence techniques such as agent-based modelling, which constructs the future research direction of the authors.

Conclusions

The results obtained from this study indicate that RBFLN is a feasible approach for integrating environmental, topographical, demographical and socioeconomic variables for predictive mapping of VL disease.

The susceptibility map of the VL indicates that riverside, nomadic villages without health-centres located in areas suitable for the vector, are the most risky areas for VL outbreaks. According to the presence of such locations, VL is common in the marginal areas of East Azerbaijan province (southern and northern part of the study area). With this in mind, infecting southern villages in the study area should raise national concerns. In the other words, from a national perspective, VL might affect other western and central provinces of Iran (i.e. Zanjan and Kurdestan) that are close to East Azerbaijan. At the same time, the RFFLN results indicate that the marginal villages of the northern part of the study area that neighbour Armenia and Azerbaijan are highly prone for a wide infection by VL. Hence, the northern part of the study area might become highly endemic in the near future. In this regard, there should be international control measures in order to prevent prevalence of the VL disease to other countries.

Acknowledgements

The first author of this study is financed by the European Union Funding Program, Erasmus Mundus, Action 2 (EMIIV). He would like to appreciate for the funding. The authors are thankful to all persons for data collection and scientific advises about VL disease. Special grateful thanks to Prof. Mazloumi Gavgani for providing information about the VL disease. Thanks are also due to Infectious and Tropical Diseases Research Center, Tabriz, Iran for the intellectual support and providing the data for the VL disease.

References

Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M, the WHO Leishmaniasis Control Team, 2012. Leishmaniasis worldwide and global estimates of Its incidence. PLoS One 7, e35671.

Beucher A, Österholm P, Martinkauppi A, Edén P, Fröjdö S, 2013. Artificial neural network for acid sulfate soil mapping: application to the Sirppujoki river catchment area, south-western Finland. J Geochem Explor 125, 46-55.

Bhunia GS, Kesari S, Chatterjee N, Kumar Pal D, Kumar V, Ranjan A, Das. P, 2011. Incidence of visceral leishmaniasis in the Vaishali district of Bihar, India: spatial patterns and role of inland water bodies, Geospat Health 5, 205-215.

Bhunia GS, Kesari S, Jeyaram A, Kumar V, Das P, 2010. Influence of topography on the endemicity of Kala-azar: a study based on remote sensing and geographical information system. Geospat Health 4, 155-165.

Biswajeet P, Saro L, 2007. Utilization of optical remote sensing data and GIS tools for regional landslide hazard analysis using an artificial neural network model. Earth Science Frontiers 14, 143-151.

Broomhead DS, Lowe D, 1988. Multi-variable functional interpolation and adaptive networks. Complex Systems 2, 321-355.

Brown WM, Gedeon TD, Groves DI, 2003. Use of noise to augment training data: a neural network method of mineral-potential mapping regions of limited known deposit examples. Nat Resour Res 12, 141-152.

Capinha C, Gomes E, Reis E, Rocha J, Sousa CA, Rosário VE, Almeida AP, 2009. Present habitat suitability for *Anopheles atroparvus* (Diptera, Culicidae) and its coincidence with former malaria areas in mainland Portugal. Geospat Health 3, 177-187.

Castillo-Riquelme M, Chalabi Z, Lord J, Guhl F, Campbell-Lendrum D, Davies C, Fox-Rushby J, 2008. Modelling geographic variation in the cost-effectiveness of control policies for infectious vector diseases: the example of chagas disease. J Health Econ 27, 405-426.

Choi J, Oh HJ, Won JS, Lee S, 2010. Validation of an artificial neural network model for landslide susceptibility mapping. Environ Earth Sci 60, 473-483.

Correa Antonialli SA, Torres TG, Conceic A¸ Filho P, Tolezano JE, 2007. Spatial analysis of American visceral leishmaniasis in Mato Grosso do Sul State, Central Brazil. J Infect 54, 509-514.

Desjeux P, 2001. The increase of risk factors for leishmaniasis worldwide. Trans R Soc Trop Med Hyg 95, 239-243.

Edrissian GH, Hafizi A, Afshar A, Soleiman-Zadeh G, Movahed-Danesh AM, Garoussi A, 1988. An endemic focus of visceral leishmaniasis in Meshkin-Shahr, East Azerbaijan province, north-west part of Iran and IFA serological survey of the disease in this area. Bull Soc Pathol Exot Filiales 81, 238-248.

Elnaiem DEA, Schorscher J, Bendall A, Obsomer V, Osman ME, Mekkawi AM, Connor SJ, Ashford RW, Thomson MC, 2003. Risk mapping of visceral leishmaniasis: the role of local variation in rainfall and altitude on the presence and incidence of kala-azar in eastern Sudan. Am J Trop Med Hyg 68, 10-17.

Fallah E, Farshchian M, 2009. Seroepidemiology study of visceral leishmaniasis among human in Azarshahr areas, East Azerbaijan, Islamic Republic of Iran. Internet Journal of Parasitic Diseases 4, 2.

Fernández MS, Salomon OS, Cavia R, Perez AA, Acardi SA, Guccione JD, 2010. *Lutzomyia longipalpis* spatial distribution and association with environmental variables in an urban focus of visceral leishmaniasis, Misiones, Argentina. Acta Trop 114, 81-87.

Fu Q, Li SZ, Wu WP, Hou YY, Zhang S, Feng Y, Zhang LP, Tang LH, 2013. Endemic characteristics of infantile visceral leishmaniasis in the People's Republic of China. Parasit Vectors 6, 143.

Hammertrom D, 1993. Neural networks at work. IEEE Spectrum 30, 26-32.

Haykin S, 1998. Neural networks: a comprehensive foundation (second edition). Prentice-Hall, Upper Saddle River, NJ, USA.

Khanmohammadi M, Fallah E, Rahbari S, Sohrabi I, Farshchian M, Hamzavi F, Mohammadpour A, 2010. Study on seroprevalence of canine visceral leishmaniasis (CVL) in ownership dogs of Sarab, East Azerbaijan, province, northwest of Iran with indirect immuno fluorescence antibody test (IFAT) and its health importance in 2008-2009. J Animal Vet Adv 9, 139-143.

Kiang R, Adimi F, Soika V, Nigro J, Singhasivanon P, Sirichaisinthop J, Leemingsawat S, Apiwathnasorn C, Looareesuwan S, 2006. Meteorological, environmental remote sensing and neural network analysis of the epidemiology of malaria transmission in Thailand. Geospat Health 1, 71-84.

Looney C, 2002. Radial basis functional link nets and fuzzy reasoning. Neurocomputing 48, 489-509.

Matouq M, El-Hasan T, Al-Bilbisi H, Abdelhadi M, Hindiyeh M, Eslamian S, Duheisat S, 2013. The climate change implication on Jordan: a case study using GIS and artificial neural networks for weather forecasting. J Taibah University Sci 7, 44-55.

Masters T, 1993. Practical neural network recipes in C++. New York: Academic Press.

Ministry of Health-East Azerbaijan province of Iran, 2006. Reported cases of leishmaniasis in East Azerbaijan province.

Ministry of Health Iran, 2008. Records of the health and communicable diseases.

Mirsamadi N, Mohebali M, Attari MR, Edrissian GH, 2003. Serological survey on visceral leishmaniasis (Kala-azar) in Azar-Shar, Azarbaijan province, northwest of Iran. Hakim Res J 6, 17-22.

Mohebali M, 2012. Epidemiological status of visceral leishmaniasis in Iran: experiences and review of literature. J Clinic Exp Path S3, 3.

Mohebali M, Javadian E, Yaghoobi-Ershadi MR, Akhavan AA, Hajjaran H, Abaei MR, 2004. Characterization of *Leishmania* infection in rodents from endemic areas of the Islamic Republic of Iran. East Mediterr Health J 10, 591-599.

Mohebali M, Motazedian MH, Parsa F, Hajjaran H, 2002. Identification of *Leishmania* species from different parts of Iran using a random amplified polymorphic DNA in human, animal reservoirs and vectors. Med J Islamic Republic Iran 15, 243-246.

Moody J, Darken C, 1989. Fast learning in networks of locally-tuned processing units. Neural Comput 1, 281-294.

Moshfe AA, Mohebali M, Edrissian GH, Zarei Z, Akhoundi B, Kazemi B, Jamshidi SH, Mahmoodi M, 2008. Seroepidemiological study on canine visceral leishmaniasis in Meshkin-Shahr district, Ardebil province, northwest of Iran during 2006-2007. Iran J Parasitol 3, 1-10.

Moustris KP, Douros K, Nastos PT, Larissi JK, Anthracopoulos MB, Paliatsos AG, Priftis KN, 2012. Seven days ahead forecasting of childhood asthma admissions using artificial neural networks in Athens, Greece. Int J Environ Health Res 22, 93-104.

Nykanen V, 2008. Radial basis functional link nets used as a prospectivity mapping tool for orogenic gold deposits within the central lapland greenstone belt, Northern Fennoscandian Shield. Nat Res 17, 29-48.

Obuchowski NA, 2003. Receiver operating characteristic curves and their use in radiology. Radiology 229, 3-8.

Palit A, Bhattacharya SK, Kundu SN, 2005. Host preference of *Phlebotomus argentipes* and *Phlebotomus papatasi* in different biotopes of West Bengal, India. Int J Environ Health Res 15, 449-454.

Peterson AT, Shaw J, 2003. *Lutzomyia* vectors of cutaneous leishmaniasis in southern Brazil: ecological niche models, predicted geographic distributions and climate change effects. Int J Parasitol 33, 919-931.

Pezeshki ZM, Tafazzoli-Shadpour A, Mansourian B. Eshrati E, Omidi, Nejadqoli I, 2012. Model of cholera dissemination using geographic information systems and fuzzy clustering means: case study, Chabahar, Iran. Public Health 126, 881-887.

Pradhan B, 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci 51, 350-365.

Rajabi M, Mansourian A, Bazmani A, 2012. Susceptibility mapping of visceral leishmaniasis based on fuzzy modelling and group decision-making methods. Geospat Health 7, 37-50.

Rinaldi L, Musella V, Biggeri A, Cringoli C, 2006. New insights into the application of geographical information systems and remote sensing in veterinary parasitology. Geospat Health 1, 33-47.

Salahi-Moghaddam A, Mohebali M, Moshfae A, Habibi M, Zarei Z, 2010. Ecological study and risk mapping of visceral leishmaniasis in an endemic area of Iran based on a geographical information systems approach. Geospat Health 5, 71-77.

Sheets D, Mubayi A, Kojouharov, 2010. Impact of socioeconomic conditions on the incidence of visceral leishmaniasis in Bihar, India. Int J Environ Health Res 20, 415-430.

Soleimanzadeh GE, Movahhed-Danesh AM, Nadim A, 1993. Epidemiological aspects of kala-azar in Meshkin- Shahr, Iran: human infection. Bull World Health Organ 71, 759-762.

Sudhakar S, Srinivas T, Palit A, Kar SK, Battacharya, SK, 2006. Mapping of risk prone areas of kala-azar (visceral leishmaniasis) in parts of Bihar State, India: an RS and GIS approach. J Vector Borne Dis 43, 115-122.

Tamook A, Moghaddam Yeganeh G, Aminisani N, Goseili F, Habibzadeh S, Sadeghi-Bazargani H, 2006. Visceral leishmanisis hospitalization in Ardebil province, Northwest of Iran. Int J Trop Med 1, 190-193.

Tsegaw T, Gadisa E, Seid A, Abera A, Teshome A, Mulugeta A, Herrero M, Argaw D, Jorge A, Aseffa A, 2013. Identification of environmental parameters and risk mapping of visceral leishmaniasis in Ethiopia by using geographical information systems and a statistical approach. Geospat Health 7, 299-308.

Yeo IA, Yee JJ, 2014. A proposal for a site location planning model of environmentally friendly urban energy supply plants using an environment and energy geographical information system (E-GIS) database (DB) and an artificial neural network (ANN). Appl Energ 119, 99-117.

Yilmaz I, 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat-Turkey). Comput Geosci 35, 1125-1138.