Modeling Skin Cancer Using Logistic Regression (Study area, Kurdistan province)

Shler Katorani*, Chia Sohrabnejad

Iranian Academic Center for Education, Culture and research (ACECR), Kurdestan, Iran.

*Corresponding Author

Abstract: *Background and Aim: Environmental factors affect the development of skin cancer. Therefore, studying the geographical location of the area has a great influence on the identification of the parameters affecting the disease. The aim of this study was to determine the spatial modeling of skin cancer and to obtain the coefficients of the effect of the variables studied in the occurrence of the disease, as well as to provide a prediction map of the probability of occurrence of the disease in the future. Materials and Methods: In this study, 1017 patients in the cancer registry system of Kordestan province were collected from the provincial medical services medical center from 2005 to the end of 2010. In the next step, maps of independent and dependent variables were prepared and then in the final stage, using the logistic regression modeling method, we determined the coefficients of the effect of independent variables on the disease. Results: In this research, the coefficients for temperature variable +0.51083534, the number of frost days +0.06045922, sun hours +0.00631544,height -0.00194397, humidity -0.16770188, Direction -0.03983796, jobs-0.21915089 land us-0.25833941, Geology -0/00201768, and for the gradient of -0.11615703. The results showed that the temperature factor, the most effective parameter, and after that, were effective in land use, occupation, average humidity, gradient, freezing days, direction, sunny hours, geology and height in the occurrence of skin cancer. Conclusion: The results of the prediction map provided by the logistic regression model indicate that in the future, the most likely occurrence of skin cancer and the region susceptible to the occurrence of skin cancer is Sanandaj, Saravabad, Marivan, bardarasha, Baneh, Kamyaran and saghez.*

Keywords: *Skin Cancer, Logistic Regression Modeling, Geographic Information System, Spatial Analysis, Kurdistan.*

INTRODUCTION

Cancer is one of the three major causes of mortality in the industrial world today. Study of any type of cancer, such as skin, stomach, breast, prostate, blood, esophagus, etc., is important from many angles. Meanwhile, the high prevalence of skin cancer in recent decades and its very negative impact on adornment of people and its negative psychological and psychological consequences has made it necessary to recognize and study it. One of the most important diseases that the human skin is affected is skin cancers. These cancers are among the most common cancers in the world, and they are more prevalent, especially in countries such as United States, Europe and Australia, whose people have white and clear skin and blue and green eyes. One of the main causes of mortality in these countries is malignant melanoma, in particular. Studies in most countries show a high prevalence of skin cancer, and the number of patients is increasing day by day (Gloster and Brodland, 1996). According to the results of studies conducted in the Netherlands and the United Kingdom, this increase can be due to the increase in day-to-day activities in open environment without adequate

clothing coverage, increased travel to the sea coast , long exposure to sunlight and reduced thickness of the ozone layer. Facing sunlight in childhood and adolescence will play a major role in the emergence of skin cancer in the future (Mohammadi et al., 2011). It is estimated that 80% of exposure to sunlight occurs before the age of 21 years (Demierre et al., 2009). The ultraviolet rays of the sun are one of the most important causes of skin cancer. Countries that are tropical and tropical, those who work in the open air and under the sunshine, and people with clear skin, are more vulnerable to malignant skin lesions. Increasing the age (age), X-rays and viruses is another cause of skin cancer. (Noorbala, 2007) Also, the high light in his article in 2007 states that in our country, due to intense sunlight in most seasons, and extensive contact with farmers, livestock breeders, mountaineers and construction workers with sun rays and the use of appropriate protective When working in an open environment, such clothing and clothing should expect a high incidence of skin cancers. The survey of the prevalence of skin cancer in the country is sporadic. Epidemiologic studies of skin cancer have been performed in most parts of the world, and most of them are the prevalence of Disease Cancer. About 25% to 35% of all cancers. In the United States, about 400,000 cases of skin cancer are reported annually, accounting for nearly 30% of all recorded cancers. The prevalence of cancers The skin in the American society is about 300 new cases per year in a population of 100,000. In the UK, the incidence of skin cancers is about 100 new ones per year in the 100,000th highest figure in Australia and the Oceania continent with around 800 new cases per year for Thousands are crowds (Noorbala, 2007). In Iran, limited studies have been performed and the incidence of skin cancers has been reported in 10,000 to 15 new cases (reported in the pathology section rather than in the community) in a population of 100,000. (Asilian, Hassan and Miqdad, 1997) The incidence of skin cancers in men and women is very different due to occupational conditions and environmental factors (Noorbala, 2007). Due to studies, environmental factors affect the development of skin cancer. The study of the geographical situation of the area has a huge impact on identifying the factors and parameters that affect the disease. The prevalence of many diseases is dependent on the geographical location of the region and environmental factors play a significant role in the prevalence of many diseases. Due to the capabilities and capabilities of the GIS, all stages of analyzing and checking the factors influencing the disease in this system can be implemented and for a wide range of problems and problems associated with diseases and health can be achieved using facilities This system offers the right solutions. Statistical and spatial and geospatial analyzes can be used to solve many health issues. Considering that so far no projects have been carried out on the spatial analysis of skin cancer in the Kurdistan province, the development of plans for the spread of skin cancer and the identification of the spread of the disease in the Kurdistan province and the determination of the impact of Because of the geographical situation, Skin cancer in the province of Kurdistan helps to identify the impact of geographic areas on the disease. Therefore, the objective of this study is to determine the skin cancer's space analysis and to investigate the skin cancer's geographic distribution. By doing this research, new horizons for the prevention and planning of health care will be open to health practitioners, which will result in a higher level of health in the country.

Materials and Methods

study area

Kurdistan province with a total area of 28203 km2 is one of the western provinces of Iran, which is located adjacent to the provinces of West Azerbaijan, Zanjan, Hamedan and Kermanshah, as well as the land with Iraq, which has more than 200 km of common border with Iraq. The geographical coordinates of the province are 34 degrees 44 minutes to 36 degrees and 30 minutes north latitude and 45 degrees 31 minutes to 48 degrees and 16 minutes east. The center of Kurdistan province is Sanandaj which is located at an altitude of 1373 meters above sea level. Other cities of the province include: Saqez, Marivan, Qorveh, Baneh, Bijar, Dehgolan, Divandareh, Kamyaran and Sarvabad. This province is located on the slopes and plains of the middle Zagros hills, is limited to the provinces of West Azarbaijan and Zanjan, to the east to Hamedan and

Zanjan, to the south to Kermanshah province and to the west to the international border with Iran. Kurdistan Province is a mountainous region. The roughness of the province consists of two parts: western and eastern. A large part of Sanandaj, Marivan and its surrounding territories make up the western part of the Kurdistan. Eastern Sanandaj is surrounded by eastern ripples. This area is surrounded by a mountainous fence composed of metamorphic and sedimentary rocks. The highest mountains in this area are the Shahstani in northern Bijar, Shidati in the center, and Ali's paw between Qorveh and Saghez (Fig. 1).
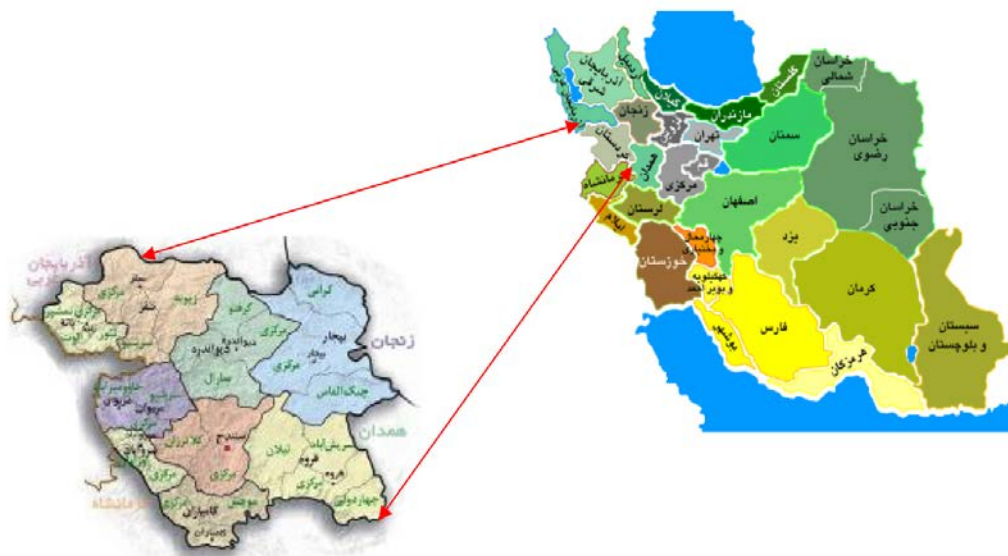

Figure 1. Location of Kurdistan Province in the country

Based on the latest administrative divisions in 1385, Kurdistan province has 9 districts, 26 departments, 84 rural districts, 1864 villages, 1765 villages and 99 vacant villages. The cities of Kurdistan province include: Saqez, Qorveh, Bijar, Sanandaj, Marivan, Baneh, Divandareh, Kamyaran, Sarvabad, which is the capital of Kurdistan province.

The variables examined

The variables studied included: Number of freezing days, Total sunny hours, Average temperature of area, Average humidity, Height, Slope, Direction, jubs, landus, Geology

In this study, the statistics and information about 1017 patients, including the patient's medical records and spatial information (address of patients including province, city, and precise postal address), were available to all patients in the cancer registry system in Kurdistan province from 2005 to the end of 2010. Kurdistan Medical Service Center was established. Then, from the governorate of Kurdistan province, the population statistics of Kurdistan province were collected from men and women, cities and districts, and the maps of Kurdistan province and the boundaries of districts and cities were prepared using the GIS format.

First, using the data of skin cancer disease in the years 2005 to 1389, a map of geographical distribution of skin cancer in Kurdistan province was provided in GIS (Fig. 2). Then, the effective parameters of skin cancer were determined, which included: average temperature, number of freezing days, total sunshine, altitude, occupation, average humidity, geology, land use, direction and tilt.

A map of all the parameters for causing skin cancer in GIS was prepared and prepared for the IDRISI medium. All maps prepared for effective parameters in the development of skin cancer disease as an independent variable and a map of the geographical distribution of skin cancer have been introduced as a dependent variable. Using logistic regression, the coefficient of effect of each parameter on the development of skin cancer disease was as shown in the table. Prognosis Skin cancer, indicating areas susceptible to skin cancer, is available.
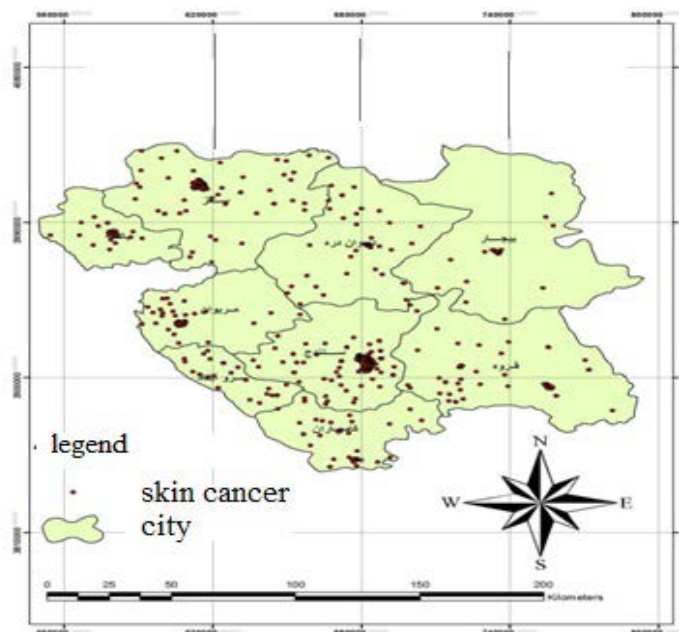
Figure 2. Map of geographical distribution of skin cancer disease in Kurdistan province

Provide independent variable mapping

In this research, the annual average of climatic factors such as temperature, number of freezing days, total sunshine hours for the whole of the province and year by year based on observations of climatological and synoptic stations of the Meteorological Organization of Kurdistan province have been calculated. The following steps are taken to generate data for entering the following analyzes:

- Average annual observations for each station were calculated.
- A point layer that represents the location of the stations was generated.
- Information on the climatic conditions in the GIS software environment was linked to the location of the stations.
- For each factor, using the IDW method, a raster layer was fitted to the mean annual for observational stations. The IDW method is based on the principle that the points that are less spaced apart are more similar.
- For each climatic condition, the estimated lattice layer with the polygonal layer overlapping the boundary of the province, and for the boundary of the province, the average values of the pixels of the rasters were calculated with the existing functions, and this amount was attributed to the spatial unit of the entire province.

For the preparation of climatological maps, due to the positioning of the synoptic and kinematic station positions, the dot layers of the stations were produced in the ARC GIS software version 9.3. Then, information about the climatic conditions in the ARC GIS software environment was correlated with the position of the stations. For each climatic factor using the IDW method and taking into account the boundary of the province, the Raster layer was prepared for the climatic factors maps.

Preparation of elevation maps, slopes, directions, geology, land use, occupations

Topographic maps were used to prepare the altitude map of the region in the form of digital elevation model. In fact, existing maps of coverage of 1/25000, which were prepared by the country's mapping organization, were used.

Disease Modeling

In order to obtain the effect of each of the independent variables studied, we developed a model for skin cancer with logistic regression method. In modeling, after gaining the importance of the independent variables studied in the occurrence of skin cancer disease, it is possible to spend time and money on the variables that have the greatest impact on the prevention and treatment of skin cancer.

- **Logistic regression**

Given that the dependent variable (skin cancer) is boolean. (Whether there is or not) is in fact between zero and one, and also due to the continuity of independent variables (temperature, altitude, occupations, etc.), so that the effect of each of the above factors on cancer Skin should be used in a way that is based on the possibility of later on and can predict the location of the disease in the future. Therefore, logistic regression was used in this study.

Logistic regression is a type of regression that predictor variables (independent) can be both quantitative and scale-dependent, but the dependent variable is a two-level category. These two categories are commonly referred to as membership or non-membership. This regression model is similar to regular regression, with the difference that the method of estimating the coefficients is not the same and maximizes the probability that an event occurs, instead of minimizing the error squared (which is done in normal regression). In the logistic regression, a concept called fortune is used for the value of the dependent variable. In the statistical term fortune, the probability of occurrence of an event (p) is the probability of the occurrence (1-p) of that event. The probability changes between 0 and 1. While fortune may be more than one. The key word in the logistic regression analysis is the mechanism called Logit, which is the natural logic of fortune (Rahmani and Esmaeeili, 2011).

Logistic regression is defined as:

$$\ln (p / (1\text{-}p)) = \beta. + \beta1\ X1 + \beta2\ X2 + ... + \beta nxn$$

The main advantage of modeling by logistic regression method in comparison with multivariate methods such as multiple regression analysis is that the dependent variable (skin cancer) can have only two values, one is the probability of occurrence of the disease and the other is the probability that the disease will not occur (Motavali, Esmaeili and Hosein Zade, 2010). Logistic regression is much more flexible than multiple regression, because it requires less assumptions than multiple regressions.

Findings

Logistic regression analysis was performed in Idrisi software using 100% samples. The results of logistic regression modeling are as follows.

Table 1. The coefficients derived from the logistic regression model and the statistical mean and standard deviations

Equivalent to variables	Independent variables	Impact facto	The standard deviation	Average
X0	constant number	-15.29934825		
X1	Direction	-0.03983796	86.210463	104.978290
X2	Average temperature	0.51083534	6.398364	7.779906
X3	Height	-0.0019439	915.635668	1114.196385
X4	Frosty days	0.06045922	49.186691	59.39777
X5	jobs	-0.21915089	7.330292	6.817317
X6	Average humidity	-0.16770188	24.179390	29498986
X7	Slope	-0.11615703	5.740709	5.834729

X8	Sunny hours	0.00631544	1501.452198	1832.804732
X9	Geology	-0.00201768	12.662274	13.565080
X10	Land use	-0.25833941	5.642315	6.550699

In the idrisi software for parameters (average temperature, number of freezing days, total sunshine, altitude, occupation, humidity, geology, land use, direction, slope) as independent variables and Bolin map of skin cancer cases as dependent variables , We aimed to determine the equation and the relationship between skin cancer parameters. The logistic regression equation for skin variables for independent and dependent variables with respect to the coefficients obtained in the table below is as follows:

Logit (skincancer) = -15.29934825- (0.039838) * X1 + (0.510835) * X2 - (0.001944) * X3 + (0.060459) * X4 - (0.219151) * X5 - (0.16770188) * X6- (0.116157) * X7 + 0.006315) * X8 - (0.002018) * X9 - (0.258339) * X10

In modeling using logistic regression method, the prediction model of skin cancer in Kurdistan province in the IDRISI software environment, according to the effective parameters mentioned in the table above, is presented in Figure 3.
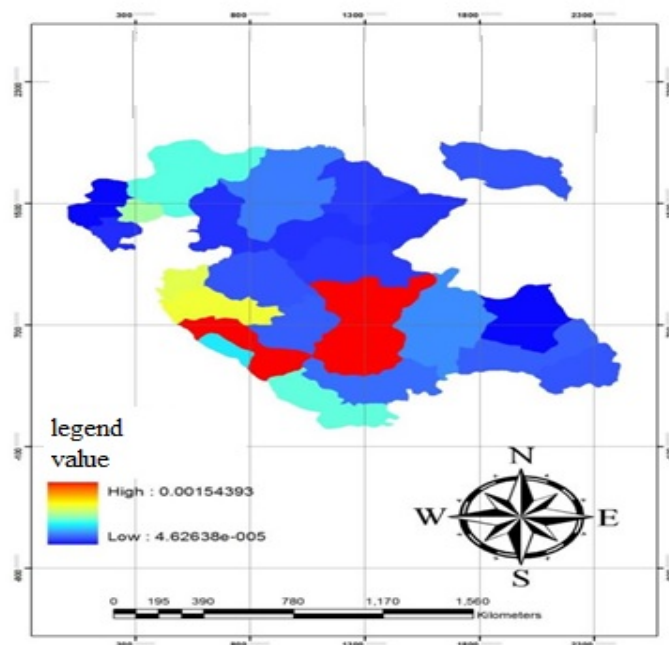

Figure 3: Modeling the Likelihood of Skin Cancer Disease by Using Logistic Regression

Discussion

The main cause of skin cancer, with the exception of colorectal cancer or melanoma, is sunlight, especially its ultraviolet component. Because between 10 and 4 o'clock in the afternoon the sun is in a more vertical position and the radiation shines more to the surface, so during these hours, it should not be exposed to naked or half-naked tan exposed to the sun. Because radiation from the sun plays a role in the development of all three skin cancers, and it is more likely to be effective in developing malignant melanoma, exposure to sunlight should be avoided. Those who work on their job to work hours of the day naked against the sun. By rubbing the anti-UV sunscreens, your body can protect your skin against both burn and cancer. Avoid

sunlight to the skin by wearing appropriate clothes, using hat caps and shading as much as possible are important ways to prevent skin cancer. Melanoma is a deadly form of skin cancer. But it can be prevented by actions of it. The high and low incidence of skin cancer in certain areas of the province indicates that health authorities in the province should pay more attention to these areas. Considering the above findings, the factor of temperature has the most effect on the occurrence of disease. Therefore, considering that the present study has been carried out in Kurdistan province, it is necessary to ensure the influence of climate parameters on the disease event, the present study in all provinces Country and country scale. The importance and impact of each of the above parameters on skin cancer can be clearly identified. In the study of skin cancer, there are no specific standards for the number of independent parameters in the occurrence of the disease. And it's best that researchers who work in the future will examine more parameters in order to have the effect of each of them on the disease.

Conclusion

The temperature factor is the most effective parameter for the onset of skin cancer. After the temperature factor, land use factors, occupations, mean humidity, slope, ice days, direction, sunny hours, geology and height are effective in the occurrence of disease, respectively.
The results of the prediction map provided by the logistic regression model indicate that in the future, the most likely occurrence of skin cancer and the region susceptible to the occurrence of skin cancer is Sanandaj, Saravabad, Marivan, bardarasha, Baneh, Kamyaran and saghez.

References

1. Asilian, As. Hassan, A. Miqdad, M. (1997). Prevalence of Skin Cancer in The Geographical Center of Iran, Research Journal of Medical Sciences. [Persian]
2. Demierre, M. F., Maguire-Eisen, M., O'connell, N., Sorenson, K., Berger, J., Williams, C., & Cabral, H. (2009). A sun protection community intervention in Quincy middle schools: Insights from the use of ultraviolet photography and its impact on sunburn. Journal of the Dermatology Nurses' Association, 1(2), 111-118.
3. Gloster Jr, H. M., & Brodland, D. G. (1996). The epidemiology of skin cancer. Dermatologic Surgery, 22(3), 217-226.
4. Mohammadi, Sa. Baghianimoghadam, M.. Noorbala, M.. Mahmoodabad Mazlomi, Sa. Falahzadeh, C. Daya, A. (2011). Given The Apparent Role In Skin Cancer Prevention Behaviors Based on Protection Motivation Theory, Skin and Beauty, 1: 70-77. [Persian]
5. Motavali S., Esmaeili R., Hosein Zade M. M., (2010). Sensitivity of Landslide Occurrence Using Logistic Regression from the Catchment Area (Mazandaran Province), Quarterly Natural Geographical, Second Year No. 5, Autumn of 2010. [Persian]
6. Noorbala, M. T. (2007). The survey of skin cancer in Yazd. Iranian Journal of Dermatology, 10, 13-19.
7. Rahmani A., Esmaeeili Gh. (2011), Performance of Neural Networks, Logistic Regression, and Discriminate Analysis in Predicting Necol, Journal of Economic Amount (The Former Economic Review). 7:152-172. [Persian].