APPLICATION OF GEOGRAPHIC INFORMATION SYSTEM (GIS) AND NEURAL NETWORKING IN IDENTIFICATION AND PREDICTION OF TOURIST HOT SPOTS: CASE OF AEGEAN REGION OF TURKEY

2021
MASTER THESIS
COMPUTER ENGINEERING

Asma RAFIQUE

Thesis Advisor
Prof. Dr. İsmail Rakıp KARAŞ

APPLICATION OF GEOGRAPHIC INFORMATION SYSTEM (GIS) AND NEURAL NETWORKING IN IDENTIFICATION AND PREDICTION OF TOURIST HOT SPOTS: CASE OF AEGEAN REGION OF TURKEY

Asma RAFIQUE

T.C.

Karabuk University

Institute of Graduate Programs

Department of Computer Engineering

Prepared as

Master Thesis

Thesis Advisor

Prof. Dr. İsmail Rakıp KARAŞ

KARABUK

January 2021

I certify that in my opinion the thesis submitted by Asma RAFIQUE titled "APPLICATION OF GEOGRAPHIC INFORMATION SYSTEM (GIS) AND NEURAL NETWORKING IN IDENTIFICATION AND PREDICTION OF TOURIST HOT SPOTS: CASE OF AEGEAN REGION OF TURKEY" is fully adequate in scope and in quality as a thesis for the degree of Master of Science.

Prof. Dr. İsmail Rakıp KARAŞ ....

Thesis Advisor, Department of Computer Engineering

Assist.Prof.Dr. Sohaib K. M. Abujayyab ....

Second Thesis Advisor, Department of Computer Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Computer Engineering as a Master of Science thesis. January 26, 2021

Examining Committee Members (Institutions)   Signature

Chairman   : Prof.Dr.Abdurrahman EYMEN (ERÜ)   ....

 Member   : Prof.Dr. İsmail Rakıp KARAŞ (KBU)   ....

Member   : Assist.Prof.Dr. Kasım ÖZACAR (KBU)   ....

The degree of Master of Science by the thesis submitted is approved by the Administrative Board of the Institute of Graduate Programs, Karabuk University.

Prof. Dr. Hasan SOLMAZ ....

Director of the Institute of Graduate Programs

"I declare that all the information within this thesis has been gathered and presented in accordance with academic regulations and ethical principles and I have according to the requirements of these regulations and principles cited all those which do not originate in this work as well."

Asma RAFIQUE

ABSTRACT

M. Sc. Thesis

APPLICATION OF GEOGRAPHIC INFORMATION SYSTEM (GIS) AND NEURAL NETWORKING IN IDENTIFICATION AND PREDICTION OF TOURIST HOT SPOTS: CASE OF AEGEAN REGION OF TURKEY

Asma RAFIQUE

Karabük University
Institute of Graduate Programs
The Department of Computer Engineering

Thesis Advisor
Prof. Dr. İsmail Rakıp KARAŞ
January 2021, 62 pages

The exploratory data analysis tools of Geographic Information System (GIS) are found very helpful to explain the extent of spatial dependence and association of various important variables of different scientific fields of inquiry. Exploratory Spatial Data Analysis (ESDA), which is part of spatial statistics tools of GIS, and the application of Global Moran's I statistics are very important in providing visual insights of spatial data. These tools are very helpful in detection and identification of spatial patterns and hotspots of an activity or process, through the measurement of spatial autocorrelation. Similarly, the use of Neural Network has also become very useful technique in predicting and projecting various spatial parameters of interest over the space. In recent years, the application of GIS and Neural Networking has gained much popularity in tourism studies. Tourism has become leading economic activity in the world in recent decades and many developed and developing countries are shifting their economic policies towards promoting tourism. However, sustainable tourism

development also requires strong policies to minimize the environmental damages. For this purpose, new research initiatives with the use of innovative techniques are required. In this background, the present study is aimed to investigate the spatial association of domestic and inbound tourist arrivals to 123 districts of all eight provinces of the Aegean Region of Turkey. Furthermore, the present research also targeted to detect and predict the potential of future hotspots regarding tourism activities in the region. The study employs ESDA techniques of GIS and Artificial Neural Network Modelling for the analysis. District-level tourism data of the Aegean Region comprising domestic and foreign tourist arrivals during 2015-2019 was collected from the Turkish Ministry of Culture and Tourism. The data was transformed into logarithm for normalization to avoid any skewness. For the analysis, the Arc GIS, GeoDa and MATLAB programs were used. The tests of local and global spatial autocorrelation and hotspot analysis was conducted to visualize and determine spatial association and location of the tourist hotspots in the region. The results of the study revealed the concentration of tourist activities in the coastal areas of the selected regions. Meanwhile, it was found that inland districts received very little number of tourists. The results of hotspot analysis confirmed the high clusters of tourists in the coastal towns of the provinces of Izmir, Aydın, and Muğla. According to the results of ANN model, among the districts of Izmir province, the most suitable areas of potential tourism hotspots are located in the districts of Konak, Foça, Aliağa, Barnova, Narlıdere, and coastal areas of Sefrihisar, Menderes and Karaburun. In the Aydın province, Didim and Soke are found most suitable areas for future tourism development. Similarly, the coastal areas of Muğla province including Bodrum, Ula, Marmaris, Ortanca and Dalaman have great tourism development potential in the region. The study is significant in the provision of useful information regarding resource allocation to the tourism hotspots and the implication of sustainable tourism policy to better utilization of tourism potential.

Key Words: Exploratory Spatial Data Analysis, Autocorrelation, Feedforward Neural Network, Tourism, Aegean Region

Science Code: 92404, 92432

ÖZET

Yüksek Lisans Tezi

TURİZM CAZİBE MERKEZLERİNİN TANIMLANMASI VE TAHMİN EDİLMESİNE YÖNELİK COĞRAFİ BİLGİ SİSTEMİ VE YAPAY SİNİR AĞLARI UYGULAMASI: EGE BÖLGESİ ÖRNEĞİ

Asma RAFIQUE

Karabük Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı
Prof. Dr. İsmail Rakıp KARAŞ
Ocak 2021, 62 sayfa

Coğrafi Bilgi Sistemlerindeki (CBS) keşifsel veri analizi araçları, farklı alanlarındaki muhtelif değişkenlerin, konumsal bağımlılıklarını ve ilişkilerini açıklamakta çok faydalıdır. Konumsal istatistik araçlarının bir parçası olan "Exploratory Spatial Data Analysis (ESDA)" ve "Global Moran's I" analizlerinin birlikte uygulanması, mekânsal verilerin görsel olarak kavranmasında önemli bir yere sahiptir. Ayrıca bu araçlar, konumsal otokorelasyonun ölçülmesi yoluyla, bir faaliyet veya sürecin konumsal yapısının ve sıcak noktalarının / cazibe merkezlerinin saptanması ve tanımlanmasında da etkilidir. Diğer yandan, son yıllarda farklı alanlardaki konumsal parametreleri tahmin ve projekte etmede Yapay Sinir Ağları (YSA) çok faydalı bir yöntem olarak dikkat çekmektedir. CBS ve YSA uygulamalarının birlikte kullanımı ise turizm çalışmalarında popüler bir yaklaşım olarak karşımıza çıkmıştır. Turizm, son yıllarda dünyada önde gelen ekonomik faaliyetlerden biri haline gelmiştir ve birçok gelişmiş ve gelişmekte olan ülke, ekonomik politikalarını turizmi teşvik etme yönünde

geliştirmektedir. Bununla birlikte, sürdürülebilir turizm gelişimi, çevresel zararları en aza indirecek güçlü politikaları gerektirir. Bu amaçla, yenilikçi yöntemlerin kullanıldığı araştırma girişimlerine günümüzde daha fazla ihtiyaç duyulmaktadır.

Bu arka plana dayanan söz konusu çalışmada, Ege Bölgesi'ndeki 8 ile ait 123 ilçeyi ziyaret eden yerli ve yabancı turistlerin konumsal hareketliliğinin araştırılması amaçlamaktadır. Mevcut araştırma aynı zamanda, bölgedeki turizm faaliyetlerine bağlı olarak gelecekteki potansiyel cazibe merkezlerini tespit ve tahmin etmeyi hedeflemektedir. Çalışmada, analiz için CBS ve YSA Modellemesinin ESDA teknikleri kullanılmıştır. Yerli ve yabancı turistlerin 2015-2019 dönemindeki hareketliliğini içeren Ege Bölgesi'nin ilçe düzeyindeki turizm verileri, Kültür ve Turizm Bakanlığı'ndan temin edilmiştir. Herhangi bir tutarsızlığa neden olmamak için veriler logaritmik hale dönüştürülmüş ve normalize edilmiştir. Analiz için ArcGIS, GeoDa ve MATLAB programları kullanılmıştır. Turistik cazibe merkezleri arasındaki konumsal ilişkiyi görselleştirmek ve tanımlamak için, yerel ve küresel seviyede konumsal otokorelasyon testleri yapılmıştır. Çalışmanın sonuçları, seçilen alanda kıyı bölgelerindeki turistik faaliyetlerin yoğunluğunu ortaya koymuş, iç kesimlerin çok az turist çektiği tespit edilmiştir. Cazibe merkezlerine ilişkin analizlerin sonuçları, İzmir, Aydın ve Muğla illerinin kıyı kentlerindeki yüksek turist kümesini doğrulamıştır. YSA modelinin sonuçlarına göre, İzmir'in ilçeleri arasında Konak, Foça, Aliağa, Bornova, Narlıdere ilçeleri ile Seferihisar, Menderes ve Karaburun kıyı bölgeleri potansiyel cazibe merkezleri olarak göze çarpmaktadır. Aydın ilinde ise, Didim ve Söke gelecekteki turizm gelişimi için en uygun bölgelerdir. Benzer şekilde Muğla ilinin Bodrum, Ula, Marmaris, Ortanca ve Dalaman gibi kıyı kesimleri de büyük bir turizm geliştirme potansiyeline sahiptir.

Bu çalışmada, potansiyel cazibe merkezleri için kaynak tahsisi yapılması ve turizm potansiyelini etkin olarak kullanmayı sağlayacak sürdürülebilir bir turizm politikası geliştirilmesi yönünde anlamlı bilgiler sunulmuştur.

Anahtar Kelimeler: Keşifsel Konumsal Veri Analizi, Otokorelasyon, İleri Beslemeli Sinir Ağı, Turizm, Ege Bölgesi

Bilim Kodu: 92404, 92432

ACKNOWLEDGEMENTS

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

SYMBOLS AND ABBREVITIONS INDEX

GIS   : Geographic Information System

ESDA : Exploratory Spatial Data Analysis

ANN   : Artificial Neural Network

LISA : Local Indicator of Spatial Association

AI   : Artificial Intelligence

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Recent developments in the computer technologies have provided great opportunities for advanced analytic analyses in various scientific disciplines. The application of such technologies in efficient data analysis with reduced cost and time has become norm of modern-day globalized world. Geographic Information System (GIS) refers to a multicomponent computer environment which is used for the organization, management, and analysis of spatial data to answer research questions of varied scientific disciplines. A GIS software enables us to locate, visualize and interpret the spatial data which is useful for understanding the relationships, trends, and patterns of different variables of interest. This state of art computer program can store and manage vast amount of data for multipurpose uses in planning and management of different spatial entities that are crucial for sustainable future of our earth. Similarly, Artificial Neural Network (ANN) technique is also important in recognition and classification of patterns in spatial data and predicting the future trends very accurately. The ANN works like the biological neurons to evaluate the data and solve the real-world problem in much efficient way.

Tourism has gained much popularity as an alternative industry in recent decades due to its ever-increasing share in worlds economy. Many developing countries of the world have made tourism an important part of their economic policies as an important source of foreign exchange earnings and diversification of economic activities. According to recent estimates, tourism has significant direct and indirect contribution to Gross Domestic Product (GDP) and job creation in the world. Turkey, a developing country, is listed among the top tourist destination of the world with a growing tourism industry. The government of Turkey implemented tourism incentive law in 1980s for maximizing the country's tourism potential in order to generate large revenues (Göymen, 2000). According to current statistics, the country ranks 6th in the list of

1

most famous global inbound tourist destination, which attracted more than 45 million tourist arrivals in 2019 contributing 41.1 billion $ tourism receipts in the national economy (WTTC, 2020). There are many geographical, historical, and cultural factors that make the country an attractive destination for inbound and domestic tourists. The unique location of the country bridging the continents of Asia and Europe, the extended coastal areas with warm climate provide great deal of sun, sand, and sea (3S) tourism.

The combination of GIS and ANN offers the chance to utilize the powerful functions of both techniques that can enhance the accuracy of decision-making and development of strategies for many important sectors of human societies including tourism. The use of spatial analytic techniques combined with ANN approach has become popular in tourism research because of the industry's strong spatial dependence, though application of ANN is still limited to specific dimensions. It is evident that global tourism industry plays vital role in local economic development by providing means of exchange earnings, income diversification as well environmental sustainability by offering alternatives to reduce dependence on natural resources (Brida, Cortes-Jimenez, & Pulina, 2016; Martins, Gan, & Ferreira-Lopes, 2017).

Many developing countries have developed national economic policies with special emphasis to tourism as cheaper source of development compared with large scale industrialization which require huge capital investments and technological advancements. Therefore, a continues growth in tourism sector is observed at global scale with development of new tourist destinations and infrastructure building to satisfy the growing demands of inbound tourists. However, economic-oriented approach in the growth of tourism may cause neglect in environmental protection and conservation which can be lethal to future sustainability. The application of GIS and ANN can provide great assistance to policy formulation and better solution to resource efficiency in this sector because of their ability to collect, manipulate and organize large amount of spatial data in lesser time.

1.2 SCOPE OF THE STUDY

The rapid growth of tourism in the world is associated to many factors like increase in leisure time, growth in income, and advancements in communication and transportation. At policy level, however, the economic value of tourism makes it prominent over many other sectors of a country's economy. Therefore, governments of many developed and developing countries of the world pay special attention to expand the tourism industry for economic growth and generating foreign revenues. In particular, the developing countries utilize the potential of tourism as an alternative to industrialization to meet the needs of growing populations and poverty reduction. However, like many other industries, the economic benefits of tourism growth are also directly based on an areas' natural resources. Tourism development can put grave consequences on the natural environment of the destination areas that may lead to irreversible damages to the physical environment. Therefore, a very cautious approach is required to make balance between economic development and environmental sustainability during the establishment of tourist destinations in an area.

Modern computer technologies like GIS and ANN can certainly play important role in better management and planning of the tourist destination to maximize the potential without compromising the natural and social environment. The use of such techniques can help providing detailed insights of spatial characteristics of a region which are necessary for the development of tourism infrastructure. Besides, the spatial statistics tools of the GIS also enable us to measure the homogeneity and heterogeneity of different locations in terms of tourism growth that is an important parameter to measure the spatial changes in tourism activities over time. Artificial Neural Network (ANN) technique, on the other hand, can help in classification and predicting of irregular tourism data very accurately for efficient management of tourism resources. The integration of both GIS and ANN is therefore very useful in locating the tourist hotspots, visualization and mapping the clusters of tourism activity, depicting the changes in spatial patterns of tourism destination, and predicting the future destinations.

The spatial patterns of tourism and location of hotspots are considered as an important indicator of measurement of tourist activity in a region. Exploratory spatial data analysis techniques identify the areas of highest and lowest tourist activities which is very crucial for resource allocation and minimizing disparities among various regions. ANN technique also provides important information regarding possibilities of new tourist destinations to overcome the issue of over-crowdedness and sustainability in the future. A sound tourism policy must deal with the problems associated to tourism growth regarding over-crowdedness, traffic congestions, capacities of accommodation and pollution. In this regard, the application of GIS and ANN can contribute to a great extent in the formulation of robust policy for management and planning of tourism. The present study has taken the case study of important tourist regions in Turkey for the application of GIS and ANN in determination and prediction of tourism hotspots.

1.3 PROBLEM STATEMENT

The efficient management and planning of tourism yield better and longstanding performance of tourist destinations which can further strengthen the sustainability of the economy. However, tourism development in an area requires a much comprehensive approach due to the involvement of complex socioeconomic and natural indicators. An unplanned growth of tourism in an area can led to various social, economic, and environmental stresses to the destinations. There are many popular tourist destinations in the world that are facing serious problems of environmental degradation, socioeconomic setbacks, pollution, and resource depletion. The problem is even worse in the developing countries which have not enough technological developments to hinder the environmental damages caused by tourism. At this point, computer-aided programs like GIS and ANN make the task of decision-making and policy formulation more organized and fruitful. The ability of these computer technologies to handle big data in short time make them very useful to locate the trends and patterns of tourism activity in an area and suggest alternative destinations for future tourism growth.

The notion of sustainable development focuses on the optimum utilization of limited natural resources of our world without compromising the needs of future generations.

In this context, the excessive exploitation of natural resources is considered a threat for the future of humanity and a race for the search of alternative sources of development is started. Among many other alternatives to minimize the dependence on natural resources, tourism is appreciated well in its practicability and feasibility in diversification of economic activities, job creation and improvements in income. However, economic development has become most attractive impetus for the tourism led growth in many developing countries where important tourist destinations are being utilized over their carrying capacities. Such practices put significant negative impacts on the sustainability of the sector which require strong policy implications for the proper management of old destinations and suitability of new areas as future destinations.

Another important point in the discussion of sustainable tourism development relates to the creating balance among three pillars of sustainability that are economic development, social development, and environmental protection. Most of the developing countries pay much attention on the economic benefits of tourism while environment remains at second priority. This situation can lead to tradeoffs between the three segments of sustainable development which further leads to serious environmental and economic damages in the society. The identification and development of new tourist destination is thus play important strategical role to overcome the tourism related environmental pressure in a region of natural and scenic attraction. The spatial statistics techniques within GIS and feedforward neural network models within ANN approach provide ample opportunities in visualization the present status of tourist clusters and predicting the most suitable areas as alternative destination to diversify the concentration of tourists. Therefore, the application of such techniques can contribute to better policy development for sustainable tourism development.

Turkey is one of the most famous tourist destinations in the world which attract millions of tourists each year. There are strong regional disparities regarding tourism development in Turkey which show large concentrations of inbound tourists in the western and southwestern regions of the country. These variations are embedded in many historical, cultural, socioeconomic and geographic factors and are very common in other regions of the world. However, growth of tourism in a region with already

exceeded carrying capacities can cause serious environmental problems threatening the sustainable economic development in the country (Tosun, 2001). Similarly, tourism development limited to certain regions of a country may also widen the regional inequalities which may lead to further problems (Khan, 2018). The review of literature has revealed some empirical studies dealing the economic and financial impacts of tourism in the country by employing different statistical and econometric approaches. However, there is a lack of studies which focus on the spatial dimensions of tourism creating a literature gap in the field.

1.4 RESEARCH OBJECTIVES

There are two primary aims of the present research. First, the present study is aimed to determine the spatial dependence of inbound and domestic tourist arrivals in the 123 districts of the Aegean Region of Turkey during 2015 to 2019. For this purpose, exploratory data analysis techniques have been employed to identify tourism hotspots in the region and depicting the spatial variations of tourist clusters over the selected period. Second, the study also seeks to predict potential areas for future tourism activities based on previous hotspot data and further independent factors of altitude, land use, nearness to roads, nearness to water bodies, geology, slope, aspect, presence of geological faults and nearness to coastline. Artificial neural networking model is used to achieve the second objective of the study.

The aims of the study are followed by the following specific objectives:

- Identification of trends and patterns of the tourist activity in the region by application of spatial autocorrelation technique (Moran's I statistics).
- Identification of tourist hot spots and detection of the change over time by employing Hot spot analysis (Getis-Ord i G∗).
- Providing visualization of similarities and differences among different tourist destinations through mapping clusters and scatterplot.
- Prediction of suitable future hotspot areas of tourism activity in the region by employing feedforward neural network model.

1.5 SIGNIFICANCE OF THE STUDY

A growing trend of tourism in the world is well appreciated for its significant share in economic growth. However, the associated environmental costs must also be determined while maintaining the old tourism destinations and planning for the new ones. Inadequate planning and management of tourist destinations have found to be destructive for the natural environment of the region in terms of increased pollution, resource depletion and environmental degradation. The situation becomes more worse in the developing countries of the world where large tradeoffs between economic development and environmental conservation are observed. Therefore, a strong tourism policy is required for maximizing the economic benefits of tourism while protecting the natural environment. For this purpose, detailed information about the spatial extent, dependence and association of tourism is compulsory to build robust tourism policy.

The findings of present study will provide useful information about the spatial patterns of tourist flows in the Aegean Region of Turkey by means of spatial techniques embedded in GIS and ANN. The identification of the currents spatial patterns of tourist flows and change over time will help resource allocation for tourism development in the region. The detailed visualization of the hotspots will also enable to policy makers to tackle the future problems of over crowdedness in some destinations. In addition, the depiction of change in inbound and domestic tourism clusters will further provide important information about the tourist movements and dispersal in the region which will help to improve service provision. Besides, the study also predicts the possible locations for future tourist destinations which will help reducing the environmental pressure in the regions with exceeded carrying capacity. The review of related literature has shown that many tourism-related studies often ignore the spatial aspect of tourism which is an important part of the sector. Therefore, the present study tries to fill the literature gap in tourism studies by providing useful information about the spatial dependence of tourism in the Aegean Sea Region of Turkey.

1.6 OVERVIEW OF METHODOLOGY

The present research has employed Geographic Information System (GIS) techniques of spatial statistics and exploratory spatial data analysis to inbound and domestic tourist data of 123 districts of the selected region. Tourism data is collected from Turkish Ministry of Culture and Tourism. Spatial Autocorrelation technique is used to determine the global spatial association of tourist flows in the region by calculating Moran's I value for a period from 2015-2019. Meanwhile, Local Indicator of Spatial Association (LISA) is used to identify the locations of tourist hotspots in the form of Moran's significance maps. The resultant maps provide the hotspot data which is further analyzed by using feedforward neural network model to predict the future tourism hotspots in the region. Figure 1.1 describes the brief methodological flow of the study. The details of applied methodology are provided in Chapter 3.

1.7 ORGANIZATION OF THESIS

The current study is organized and structured as follows. After the introduction, Chapter 2 provides the basic concepts of GIS and Artificial Neural Networking along with the detailed review of related studies. The chapter discusses the use of GIS and ANN in the field of tourism and important findings of previous studies as well the literature gap which is filled by the present study. Chapter 3 describes the details of methods applied in the study and the processes involved in making digital maps of the area to fulfill the objectives of the study. Chapter 4 explains the findings of the study and detailed discussion. Following the results and discussion, Chapter 5 includes the concluding remarks with some recommendations and policy implications regarding tourism strategies in the region.

```
                    Selection of Study Area

                    Data Colllection (Tourist Arrivals)

                    Data Processing in
                    GIS

    Spatial Statistcis      Exploratory Spatial Data      Hotspot Analysis
                                  Analysis                  (Getis-ord)

                          Derived Maps

                          Road        Water
                        Nearness     Nearness      Slope      Geology

    Tourism Hotspots
                          Faults     Coastline     Aspect     Landuse
                                     Nearness

    Dependent Variable              Explanatory Factors

         Development and Training of Feedforward Neural Network Model

                          Map of Predicted Hotspots
```
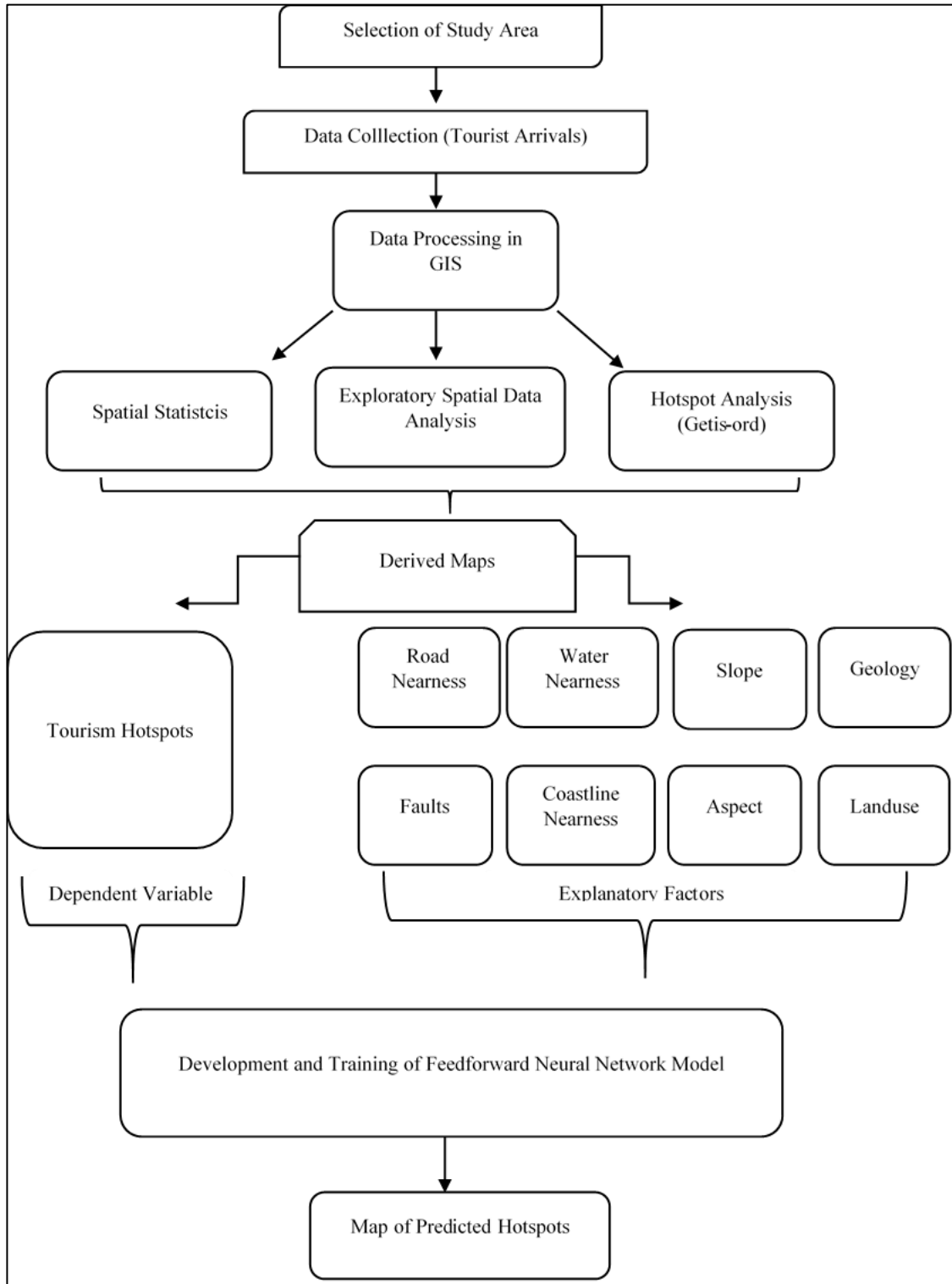
Figure 1.1. Overview of Methodology.

9

CHAPTER 2

LITERATURE REVIEW

A comprehensive literature is found on the application of different GIS and ANN techniques in much diverse fields of scientific disciplines. This chapter describes the basic concepts of spatial analysis and machine learning briefly, followed by a detailed review of literature on the use of different methods of GIS and Artificial Intelligence in tourism studies. Besides, the chapter also presents the important findings of various studies on tourism using spatial statistics and ESDA alongside different neural network models. In the last part, the chapter also detect literature gap which would be filled by the present study.

2.1 SPATIAL ANALYSIS IN GIS

A Geographic Information System (GIS) is a computer-based system which comprises tools and techniques for storing, analyzing, manipulating, organizing, and visualizing the spatial data in short period of time. The program has great ability to deal large amount of quantitative data associated to any location on earth in digital environment to create digitized maps, building models and projecting real time processes in visualize form. According to the Environmental Systems Research Institute (ESRI), an international provider of GIS software, "GIS is an organized collection of computer hardware, software, geographic data and personnel designed to efficiently capture, store, update, manipulate, analyze and display geographically referenced information". The data can be presented in vector (points, lines, and polygons) and raster (imageries) forms as described in Figure 2.1.

Spatial analysis in GIS enables us to transform the raw data about different attributes of an area into useful information. The identification and determination of patterns and trends of various variables of interest helps to understand the complex spatial interactions among them which is necessary for adequate decision making. Many scientific fields including, geography, health sciences, ecology, architecture, management studies, business sciences etc. have been utilizing the GIS spatial analysis techniques as vital part of innovative research.
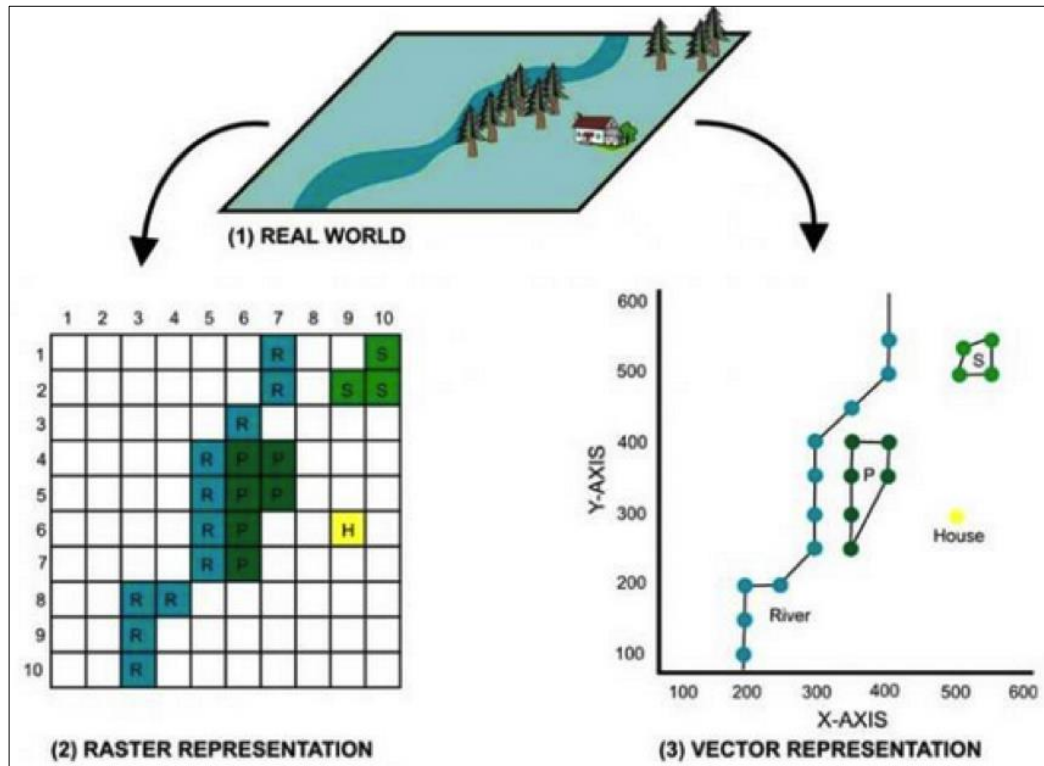
Figure 2.1. Presentation of Raster and Vector Dataset. (Source: ESRI)

2.2 ARTIFICIAL NEURAL NETWORK MODELLING

Artificial Neural Networks are very useful part of machine learning techniques which function like a learning mechanism in an organism through nervous system to respond an external stimulus. The nervous system of humans and many other organisms consists of various cells known as neurons which are connected to each other to transmit signals of various stimuli to different parts of body (Figure 2.2a). The idea of artificial neural networks is based on simulation of biological nervous system in which some computation units work as neurons connected to each other with weights of input information (Figure 2.2b). In this way, an artificial neural network measures the function of inputs by transmitting the measured values of input neurons to output neurons employing the intermediate factor of weights.

For the purpose of research, artificial neural networks are trained through different input-output pairs of learning functions which can be further used to predict the outputs. The technique has become much popular in prediction of many spatiotemporal

11

processes including flood site predictions, landslide susceptibility, agroforest zoning and tourism forecasting etc.
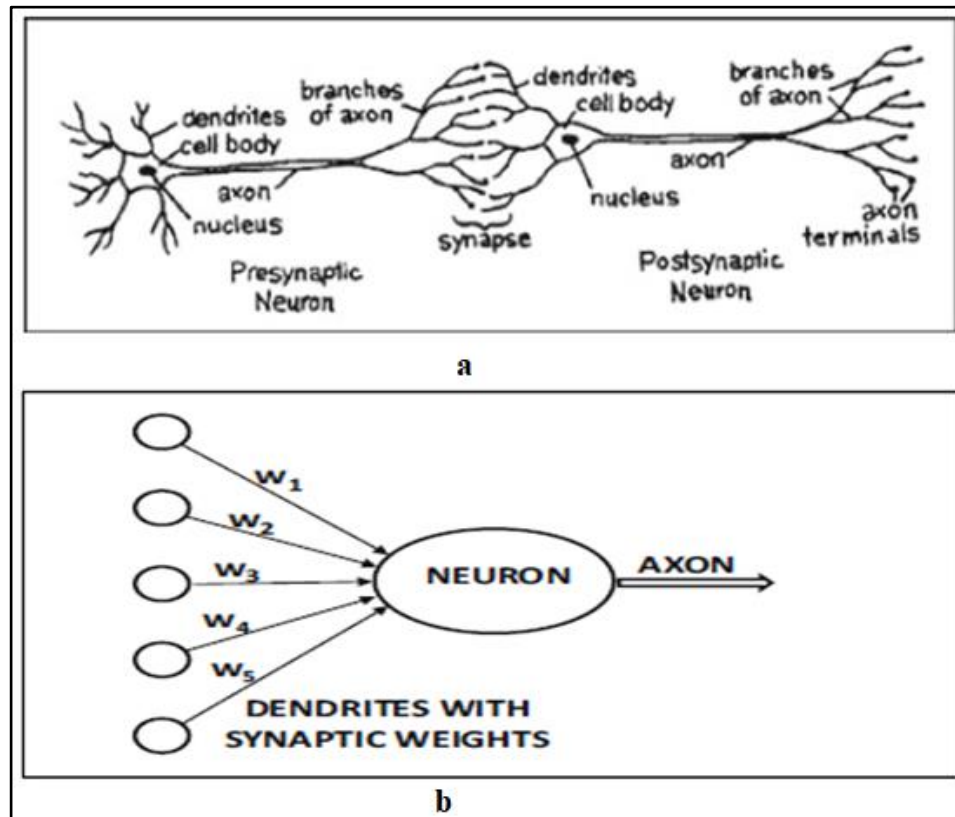
Figure 2.2. (a) Biological neural network, (b) Artificial neural network. (Source: Aggarwal, 2018)

A model based on artificial neural network have at least one or more than one layers of different neurons. Many artificial neural network models like Perceptron, Linear Associator, Multi-layer feed-forward networks with Back-Propagation (BP) learning, the Boltzmann machine and the Grossberg model use the output from the units from one layer which can perform the function to activate neurons in only the next layer. There are several others model like Kohonen nets and Hopfield model in which the neurons of same layer can also be activated by the signal. Some other models such as Self- organizing Feature Map (SOFM), offers the possibility in which the network connects a vector of inputs to a two-dimensional grid of output neurons. Different kinds of neural network models are shown in Figure 2.3.
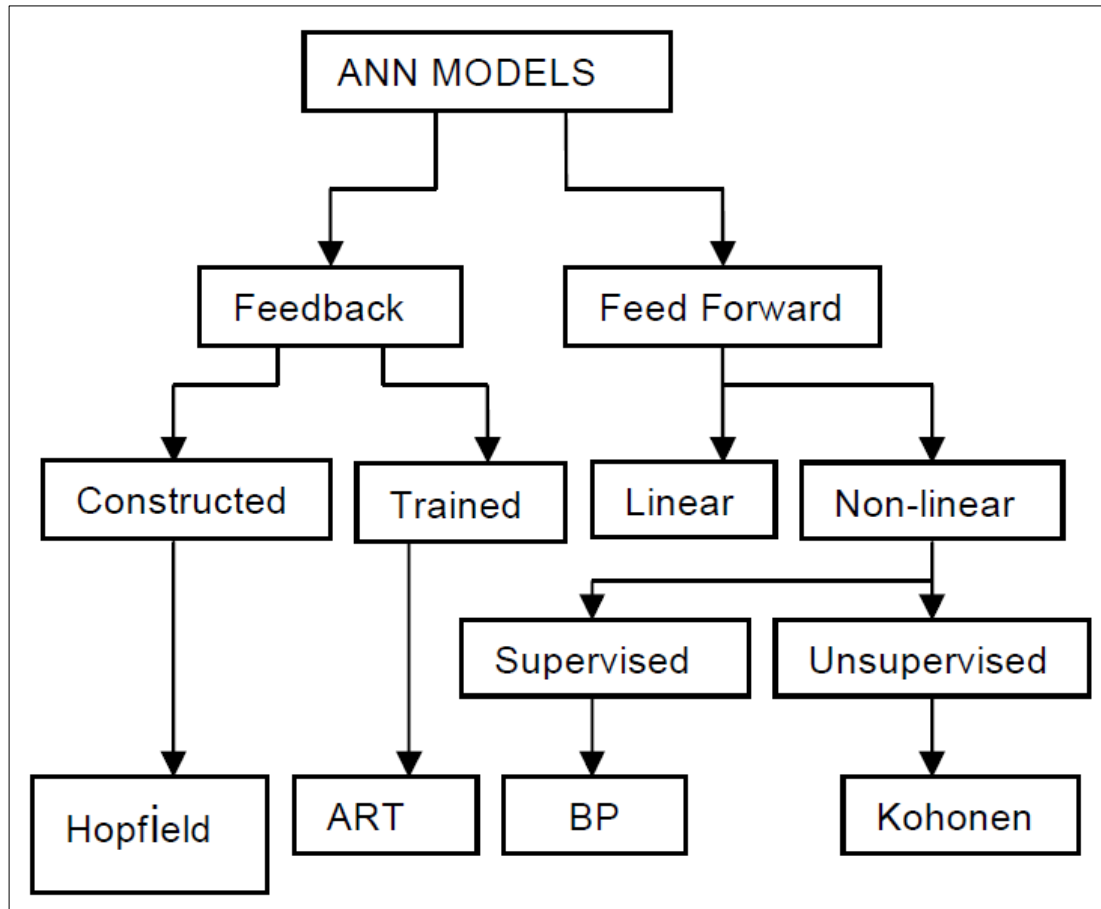
Figure 2.3. General Classification of Neural Network Models. (Source: Malik, 2005)

2.3 APPLICATION OF GIS IN TOURISM RESEARCH

Geographic Information System (GIS) provides plenty of sophisticated tools to management, organization, and planning of tourism industry all over the world. This computer-based system offers variety of functions integrated with other components of information technologies to develop digital maps, visualization of places on internet and interactive maps which play critical role in tourism planning and advertisement. Thus, GIS helps the development of tourism destination by handling big tourism data and integrating it into computer environment for further complex analysis. GIS tools can contribute to tourism planning at great extent by identifying different components of tourism infrastructure like transportation, accommodation, and attractive places in

geometric, thematic, and topologic manner. As tourism planning requires integrated approach, GIS can facilitate to build useful models which are based on real world experience.

2.3.1 Use of GIS in Tourism Planning and Management

Recent research in different fields of society such as ecology, forestry, economics, geography, and environmental sciences has employed the site suitability analysis of GIS for solving many spatial problems. The powerful tools of spatial analysis are applied not only in selection of suitable locations for large number of socioeconomic activities but also for urban and regional planning. As part of tourism planning, land use suitability analysis is used for detailed insights into the natural and physical properties of area which are necessary for building tourism related infrastructure. There is plethora of empirical studies employing suitability analysis in tourism to identification of best location for tourism development in different regions of the world. GIS approach is tourism studies is also favored by the advocates of sustainable tourism development to measure, visualize, and predict the status of sustainability in any tourism destination.

Site allocation for various types of tourism through GIS offers great opportunities for efficient consumption of tourism potential in an area by maximizing the economic benefits for the local communities. As mentioned earlier, GIS has great ability to transform the real-world data into different layers of vector data which further helps understanding the complex spatial relationship among different variables. Map layers of different natural, physical, socioeconomic, and environmental parameters provide the bases for selecting most adequate areas for future tourism development which would last for long without environmental damages. Spatial analysis of various important features is also referred to multicriterial decision analysis which employs weighted sets of various spatial factors to reach accurate conclusions for best suited sites of tourism. Such methods are highly recommended through results of many empirical studies for better planning and management of tourist destinations.

Large amount of literature on tourism planning and management has utilized the GIS based techniques of multicriteria decision analysis combined with other methods such

as analytic hierarchy process and spatial regression to identify and detect the possible sites for alternative tourism activities in a tourist region. In this way, tourist activities may be diversified and dispersed to different areas which reduce the over crowdedness and environmental pressures on the tourism concentrated sites of a destination. Besides, site suitability for developing new tourist destinations is also analyzed to enhance the tourism related development in other regions of a country to reduce the regional disparities within a country. Some selected studies have been discussed in following paragraphs to present the use of GIS in locating tourism potential and planning the destinations.

In a recent study from Iran, Eftekhari and Mahdavi (2019) GIS techniques of multicriteria decision analysis has been used for determination of potential of nature-tourism in Tehran. For the purpose of analysis, GIS based multicriteria technique is employed based on three criteria of Site Feature, Infrastructures and Services, and Risk and Sensitive Areas. Further, the map of areas with higher to lower suitability were produced to depict the important locations of tourism activities. The results of the study also produced the statistical measurement of the extent of the suitable land for tourism development that was approximately 15.87 % of total studied area. Hence, the study provided practical implications for strong tourism policy in the region.

Similarly, an empirical study in China by Deng et al. (2019) utilized sophisticated tools of GIS multicriteria analysis for the selection of appropriate sites for ski tourism combined with weighted linear regression. Many important factors of physical environment like topography and climate conditions integrated with multiple socio-economic indicators were analyzed with help of GIS techniques. Suitable sites for the development of ski resorts were determined by generating metrices of selected factors. The results of analysis confirmed the compatibility of 84% of established ski resorts to the integrated index based on several weighted spatiotemporal and socioeconomic factors. After finding suitable locations, several important strategies for the development of ski resorts were formulated to help decision makers which incorporates the future climate change and socioeconomic scenarios.

In a recent study related to GIS conducted by Pan et al. (2019), identification of suitable sites for health tourism development in mountain regions of China has been made by applying combination analytic network process and with Delphi approach and multicriteria decision analysis. In order to provide the comprehensive information regarding potential of health tourism in selected area, suitability maps of four different activities were developed with the help of GIS based on several geographical factors related to wellness tourism. Detailed maps of best locations were produced for the activities of health preserving, sports and recovery, medical healing and aged nursing in the selected districts. The study suggested some useful policy implications for better planning and management of health tourism in China.

2.3.2 Use of GIS Spatial Statistics in Tourism

The use of Exploratory Spatial Data Analysis (ESDA) has also become popular in many distinct areas of investigation as a vital tool to measure spatial variations of different variables(Blazquez and Celis 2013; Gomes, Cunto, and Silva 2017; Nawaz et al. 2019). Some recent empirical studies have employed the spatial statistics tool of GIS to explore the spatial patterns and trends of tourism activities in various parts of the world. Such studies have provided very useful information in understanding the extent of tourism activities and depict the variation over space to determine the patterns of heterogeneity and homogeneity. The findings of such studies have presented very detailed insights of spatial dependence and association of tourism to locate hubs of tourism growth and development in different regions of the world.

In a study conducted in Serbia, Stankov et al. (2017) analyzed the trends and patterns of tourist activities at municipal level by employing ESDA and autocorrelation techniques. A significant level of disparities in tourism movements was observed in the findings of the study which is not addressed in the tourism policy of the country. Moreover, the results also revealed a declining trend of domestic tourist clustering over time in the country. On the other hand, the inbound tourists showed relatively stable and growing trend of high clusters. Such changes in the distribution of tourist clusters in the region are required to be addressed properly for sustainable tourism development in the future.

The spatial statistics techniques of GIS have been applied by Chhetri et al. (2013) to analyze the clusters of tourism and hospitality employment in Victoria region of Australia. Some statistical inferences were made on the data of employment related to big tourism industry employers. Besides, the study was also aimed to explore the vicinity of tourism industries by determining the distance from the areas of high economic importance. In this regard, spatial autocorrelation techniques were employed to locate the clusters of selected variables and their patterns over the space. The results of the study revealed the presences of five important clusters of tourism employment which had the potential for developing into big hubs of industrial and business activity in the region.

Similarly, Yang and Wong (2013) used the spatial statistics techniques in GIS environment to figure out the tourism patterns and clusters in China. They evaluated the domestic and inbound tourism activities using district level data to describe the patterns in more sophisticated manner. Besides, their study also utilized the global and local Moran's I statistics as part of the ESDA to locate the global and local patterns of tourism hotspots. The findings of the study revealed the location of major clusters of tourism activity near the coastal areas and insular areas of the country. It was also found that large disparities were present in distribution of tourism growth within different regions of the country. Useful suggestions for resource allocation and policy implication were detected after the analysis.

In another study in China, Hu, Neelam, and Green (2016) applied the GIS based spatial statistics techniques to explore the activities of inbound tourists in mainland districts. Spatial Autocorrelation analysis revealed the presence of hotspots and cold spots regarding tourism patterns in the areas of study. They also applied multiple regression analysis to determine the impact of other variables on international tourist arrival and found that transportation structure and service facilities had great impact on the number of tourist arrivals. Again, hotspots of tourist activities were found near the coastal areas and adjacent areas to big districts while cold spots were located in western part of the country. The study suggested important policy implications for better planning and management of tourism activities in China.

Similar methods were used in a study from Australia in which Chhetri et al. (2017) applied spatial econometric modeling and spatial autocorrelation techniques to locate the clusters of tourism employment in rural and urban areas of the selected region. The results of spatial econometric modeling revealed the growing clusters of tourism employment in the rural areas of Australia are mostly related to operation services of tourism. In contrast the high clusters of tourism employment were found mostly associated to hospitality services of tourism. Moreover, the study explored that spatial clustering of tourism employment were influenced by six location specific factor including the availability of tourism attractions; proximity to the coast; the road density network; accessibility to employment within the Melbourne CBD; the scale of the regional economy; and the advantages and disadvantages associated with economic resources. They further argued that the clustered played important role in tourism related employment to adjacent catchment areas of tourists.

Again in a study from China, Yang et al. (2013) determined the spatiotemporal patterns of tourism growth by employing spatial statistics techniques of GIS. Strong spatial dependence of tourism activity was found in the analysis of time-based tourism data. High clusters of tourism were observed in the eastern parts while the western parts remained cold spots in terms of tourism development. The autocorrelation measurements showed a slow increase form negative value to positive one during the time. The change in the value of Moran's I statistics revealed the spatial clustering of the tourist patterns in the region over time. With useful findings of the analysis, the study made policy suggestion for robust tourism development.

Xu et al. (2020) evaluated the effect of haze pollution on international tourist arrivals by employing spatial econometric and spatial statistics techniques. The study explored the direct and spill over impacts of haze pollution in terms of heterogeneity based on quantitative analysis. It was found that both the pollution and inbound tourist arrivals had significant association of autocorrelation and spatial spillover effects. The findings of the study revealed significant direct and spillover effect of haze pollution on the growth of inbound tourism at regional level. However, the effects were found to decrease slowly from south to north. The results suggested important policy strategies for government to control haze pollution and improve tourism.

18

In another study from Iran, Talebi et al. (2019) analyzed the road netwrok used for tourism purpose in a protected area and proposed the suitability of areas for new netwrok of roads by employing Multicriteria Dicision Making approach through GIS techniques. A fuzzy logic approach was used to standardize the map layers created in GIS environment based on different criteria of physical peoperties of the area. The resulted suitability map enabled to develop 14 diiiferent scenarios of proposed road network for the purpose of access to the area for recreational activities. All the proposed scenarios were tested according to technical, environmental, and socio-economic criteria for optimal designed road network. Later, a modified netwrok of roads was proposed based on the optimal design network and compared with the actual road network. The proposed road netwrok was identified as more suitable alternative road network.

2.4 APPLICATION OF ANN IN TOURISM

It is mentioned earlier that an Artificial Neural Network consists of various neurons which are interconnected computed units structured in different layers. Input neurons are processed under different weighted mechanisms to yield the output neurons for prediction of different variables of interest. This highly sophisticated process of relationship among different nodes and neurons helps in generating knowledge which can be organized and structured to make accurate decisions (Çuhadar et al., 2014). Thus, ANN technique can be very helpful in prediction and forecasting of different spatial and non-spatial phenomenon in reliable manner. There is growing tendency of use of ANN in tourism forecasting in recent years. Many scientific research works is available on the use of ANN in prediction and forecasting of tourist arrivals. Some of the empirical studies on tourism employing ANN modelling are discussed below.

In a study on tourism demand forecasting, Wiley (2007) compared three models of neural networks to test their performance in forecasting the tourism in more accurate term. The three neural network models used in the research namely a multi-layer perceptron, a radial basis function and an Elman network. The results of the study that multi-layer perceptron and a radial basis function models of neural network provide more accurate results in tourism demand forecasting as compared to Elman network.

In tourism industry, the findings of the study contribute much by providing the useful information regarding implementation of artificial intelligence for better prediction.

Hsu and Wang (2008) applied fuzzy grey model FGM(1,N) and back-propagation neural networks (BPNN) models for predicting the demand of tourism. Their research aims at providing more detailed information regarding the vulnerability of tourism demands to various fluctuations in the industry. The study used the forecast model on the travel demand for USA and Japan from Taiwan for period 1990-2003. They also applied several benchmark univariate forecasting models including the naïve method, exponential smoothing model, Holt's method, and linear regression. The findings of their study revealed the much better performance of FGM (1,N) model of neural network for more accurate prediction as compared to benchmark statistical methods of tourism demand in the selected period. The results of their study have significant potential for the management of tourism.

Moutinho et al. (2008) determined the tourism demand for Taiwan from the mainland China by employing neural network modeling technique. They used several explanatory variables of GDP, Income, Consumer Price Index, and Exchange rate to determine the demand of tourists for a period 1999-2005. Besides, the study also employed other methods of forecasting like ARIMA for comparison with neural network model. The findings of their study confirmed the accuracy of selected neural models in tourism demand forecasting as compared to other predictive methods based on traditional statistics and computer modeling.

In another study from Taiwan, Huarng, Yu, and Parellada (2011) applied fuzzy time-series model to forecast the tourism demand and determined the fluctuations in the tourist arrivals through regime switching model. They clustered the time series data of tourism into three regimes of different number of tourist arrival which were low, middle, and high. The results of the study revealed that tourism industry in Taiwan remained prosperous for the selected period 1984-2008. Only the outbreak of disease SARS caused regime switch in Taiwan's tourism industry. Meanwhile, the outperformance of selected neural network model was confirmed for accurate prediction as compared to other methods.

It is argued that combination of different forecasting models yields better results as compared to single forecasting models. Moreover, combination of linear models provides better and accurate results and tested extensively in the literature. The non-linearity is not much investigated despite its importance in understanding the tourism demand. The performance analysis of different linear prediction models has been questioned by Cang (2011) after testing the nonlinear model for tourism forecast. In this study, the author proposed multilayer perceptron neural networks (MLPNN) on quarterly data of inbound tourist arrivals in United Kingdom to determine the future patterns of inbound tourism. The results of the study found the better performance in accurate prediction of tourism demand in both seasonal and semi-seasonal trends of tourist arrivals.

In a study conducted by Chen, Lai, and Yeh (2012), empirical mode decomposition (EMD) and neural network models have been tested for predicting tourism demand in Taiwan. EMD approach has applied for the purpose of decomposition of raw data into limited set of intrinsic mode functions (IMFs) and a residue. The study then employed back propagation neural network to determine the predictive values of tourist arrivals to Taiwan. It is found that the proposed model of combining the neural network model with EMD resulted in more accurate forecasting when compared to simple back propagation model.

In order to provide the detais of tourism performance competitiveness, Fernandes et al. (2013) proposed set of different models based on neural network techniues. The study evaluated the time series data of tourist arrivals in two different regions of Portugal. They used a feedforward structure model trained with backward propagation to predict the tourism demand for the year 2008 and 2009 based on the data for period 1987-2009. It was found that the selected model performed well in high accuracy of forescasting tourism demand in non linear fashion.the study further suggested the use of different other economic variables for future predicitons.

Claveria and Torra (2014) evaluated the forecasting performance of neural netwrok models in comparison to traditional statistical methods like ARIMA and SETAR methods in tourism demand for Spain. The study used annual data of tourism in

Catalonia region of Spain from 2001 to 2009. It was revealed that ARIMA model provided more accurate prediciton as compared to SETAR and ANN models. Besides, the study also found the better results while using total number of tourist arrivals in contrast to overnight stays. The study suggested to consider structure optimization and incorporating extra memory values to improve the performance of neural netwrok models in accurate prediciton.

In a study conducted by Saeidi et al. (2017), compared the performance of different methods in landscape aesthetic suitability mapping of Golistan region of Iran. A comparison of Multi-Criteria Evaluation (MCE), Logistic Regression (LR) and Multi-layer Perceptron (MLP) neural network was made to evaluate the accuracy in landscape aesthetic mapping performance. The spatial accuracy was determined by the use of Receiving Operator Characteristic (ROC) method. The values of ROC were found to test the high performance of selected models in landscape aesthetic mapping. It was found that Multi-Criteria Evaluation (MCE) methods with expert opinions outperformed the other methods in accurate suitability mapping.

Li and Cao (2018) compared different linear and non linear methods of tourism forecasting to find out the most suitable method for accurate prediciton. They used Long Short Term Memory Neural Network (LSTM) methods to forecast the tourism demand using the data set of tourist arrivals in important pagodas of China named Small Wild Goose Pagoda. Meanwhile, the study also performed the prediciton tests based on Auto Regressive Integrated Moving Average (ARIMA) model and Back Propagation Neural Network(BPNN) to compare with the findings of LSTM model. The results of their study revealed that LSTM model performed best in accurately predicitng the tourism demand that that of ARIMA and BPNN models. Moreover, the study also suggested to add more tourism factors like weather, tempertaure and working days for future researches.

In a recent study conducted by Gregorić and Baldigara (2020), the seasonal demand of German tourists for Crotia was caculated through implementation of augmented HEGY test of unit root and neural networks. Stationary properties of the tourist data was checked and found non-stationary in relation to zero frequency and seasonal

frequency. After confirming the non stationary properties, several artificial neural network model were proposed and tested to determine the tourism demand. The results of the study revealed that purchasing power and high service price played determinental role in the tourist behavior to visit Crotia. Further, the results revealed that GDP and price did not affect the tourist demand at significant level.

2.5 APPLICATION OF GIS AND ANN IN TOURISM STUDIES RELATED TO TURKEY

GIS techniques have become very popular in spatial analysis of tourism destinations for adequate management, planning and resource efficiency. There are various tools embedded in GIS which can facilitate desired analyses according to the nature of tasks. The demand for cruise tourism in İzmir, Turkey was determined by Çuhadar, Iclal, and Ceyda (2014) by comparing Multi-layer Perceptron (MLP), Radial Basis Function (RBF) and Generalized Regression neural network (GRNN) of neural netwroks. They used the monthly data of tourist arrivals for cruiseships in İzmir province of Turkey for the year 2014. RBF neural network model was found more suitable for high accuracy of tourism demand compared to MLP and GRNN models in the study. The study further suggested to use more robust models without problems related to structure failure and over training. In order to overcome the problems, the study also suggested the use of support vector machines (SVMs); rough sets; fuzzy logic; genetic algorithms (GA) and adaptive neuro-fuzzy inference system (ANFIS) for more accurate forescasting of tourism demand.

In an effort to provide more improved performance of neural network in seasonal time series forecasting in tourism, Hamzaçebi (2008) ANN structure to determine the number of input and output neurons. After the construction of model, it was tested on four real time tourism time series data. The findings of the study revealed that the proposed ANN model outperformed the traditional statistical methods of prediction. It was found that the proposed ANN model had minimal errors of forecasting tourism as compared to other methods.

Khan (2018) utilized the spatial statistics techniques to explore the spatial association of inbound and domestic tourist arrivals in provinces of Turkey. Tourism data from

2006 to 2016 were analyzed to measure the spatial autocorrelation statistics over the time in 81 provinces of Turkey. It was found that tourism activities had strong positive spatial association in the country. The western provinces of the country showed the high activity of tourism and thus were placed in the category of tourist hotspots. In contrast the southeastern provinces remained in the cold spots in terms of tourism development. The disparities observed in the analysis were attributed to different historical, political and most importantly physical characteristics of the regions. In addition, the study provided some useful policy suggestions in resource allocation to tourism development in priority areas. The study utilized province level data which may not depict the details of smaller spatial units.

In another study from Turkey, Kaptan Ayhan et al. (2020) applied the ELECTRE multicriteria decision analysis with GIS for determination of suitable sites for rural tourism in Yenice, Turkey. The area is divided based on natural properties including the land-use, slope, climate, geomorphology and nearness to roads and water bodies in a GIS environment. Besides, the important places of historical values have also been added in the analysis as tourist attraction to produce suitability maps of various tourist activities. After the analysis, the area has been categorized into the nine suitable locations for tourism activities in different rankings of priority. These suitability categories have been developed on three rankings of different activities of tourism including, mountaineering, trekking, and flora observation.

2.6 LITERATURE GAP

Review of the related literature has enabled us to find out the significant gap in literature, which is to be filled by the current study. In global context, most of the previous studies employing GIS have focused on the identification of tourism potential or site suitability for the development and management of tourist destinations. The findings of such studies have provided ample information on various important spatiotemporal and socioeconomic factors which play determinantal role in tourism development different areas. However, such studies have not provided detailed information about the changing patterns and trends of tourist flows which are very crucial in prediction of tourist behavior, destination images, resource allocation and management of the destination.

Very few studies have utilized the spatial statistics tools of GIS to determine the spatial association and dependence of tourist clusters and hotspots. Most of the previous studies employing exploratory spatial data analysis (ESDA) techniques have only focused on the determination of spatial patterns and none of them has performed further analysis to predict the future of tourism development in selected study areas. The present study has not only employed the spatial statistics techniques to identify the patterns and hotspots, but neural network modeling has also been applied to predict future sites of potential hotspots in the region. Therefore, the findings of the present study would provide more valid and authentic information regarding sustainable tourism development by combining both GIS and ANN techniques.

The analysis of literature has also revealed that despite enjoying the status of famous tourist destination in the world, there is lack of sufficient literature on spatial dimension of tourism in Turkey. A great deal of tourism literature in Turkey has focused on the economic growth of tourism and spatial dimension is often ignored. Many previous studies only analyze the tourism potential in different areas of the country and spatial dependence is seldom discussed. Therefore, the present study is aimed to fill the current literature gap and add significant amount of knowledge which would open ways for future research based on combination of GIS and ANN.

It is also found that application of ANN is only limited to tourism forecast and demand modelling employing time series data. There are very few studies which have employed the ANN technique in projecting the tourism potential of some areas. At present, according to the knowledge of researcher, no single study has employed ANN in identification or prediction of suitable locations for tourism in Turkey. Therefore, the present study would be pioneer in utilization of ANN in predicting tourist hotspots in Turkey.

CHAPTER 3

METHODOLOGY

This chapter provides detailed methodology involved in the present research. The present study has employed exploratory spatial data analysis (ESDA) techniques as part of spatial statistic tools provided in GIS. Spatial autocorrelation, Global Moran's I statistics, local indicator of spatial association (LISA) and Getis Ord method of clustering and hotspot analysis have been applied in determination of spatial association and identification of hotspots related to tourism activities. Feedforward Neural Network Modelling is used for the prediction of tourist hotspots based on selected indicators.

3.1 STUDY AREA

The Aegean Region is one of the major geographic regions of Turkey which lies in the western part of the country. The region derived its name from the Aegean Sea which borders in the west. In the north, lies the Marmara region, while the Mediterranean and Central Anatolian regions of Turkey are in the east and south of the Aegean region, respectively. This region has the longest coast in comparison to other coastal regions of Turkey. Administratively, the region is divided into eight provinces which are further subdivided to districts. There is total 123 districts in eight provinces of the region out of which 30 districts are included in province of İzmir, third largest province of Turkey. It is estimated that total area of the Aegean Region is approximately 90,000 km^2 (Figure 3.1).

A large part of the coastal areas of the region experiences prevailing Mediterranean climates with characteristics of hot and dry summer and mild winters. The effects of prevailing coastal climate are extended towards the inland areas up to 100-150 km, after which, it gradually predominates by continental type of climate. The area has several geological fault structures which make it earthquakes prone. Mountains run perpendicular to the shore of the Aegean Sea. Streams bend formation is common. It is a region rich in geothermal resources of the energy. The population is mostly concentrated in the coastal areas gradually dispersing the inland areas. Turkey has a

population density above the average. Cold raids begin to be seen in the northern parts of the region. While January is the coldest month in the Aegean Region, the hottest month is July. The amount of precipitation can vary between 500-999 mm. The Aegean Region itself is an alluvial plain. The region with many plains is rich in this respect compared to many other regions.
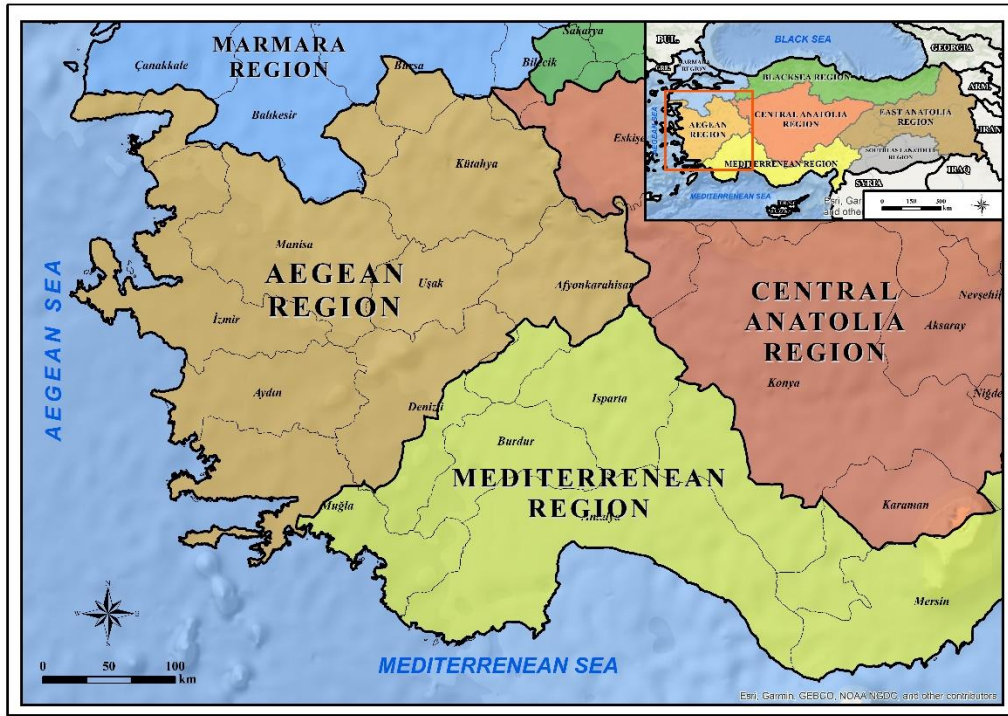

Figure 3.1: Location map of Aegean Region of Turkey. (Source: ESRI)

Maquis is the dominant type of vegetation along with some forest areas seen in the Mediterranean Region, according to the land conditions. The vegetation cover consists of scrub in sections up to 400 meters, while the vegetation seen in areas above 400 meters is forested area. The extended coastal areas of the region attract large number of tourists for leisure activities associated to Sun, Sand and Sand (3S tourism). The region is popular destination of inbound tourists as well the domestic.

3.2 COLLECTION OF DATA

The present study has used the yearly data of domestic and inbound tourist arrivals of the Aegean Region of Turkey for 2015 to 2019. The district-level data was collected from the yearly bulletins of the Turkish Ministry of Culture and Tourism

(https://yigm.ktb.gov.tr/) . The data was arranged and organized for the purpose of the study. In order to avoid any skewness problems, all the statistical data were transformed to logarithm form to yield robust results. The software of ArcGIS 10.5 and GeoDa were used for spatial analysis.

3.3 MAPPING THE DATA

The first step of spatial analysis in this research starts with the formulation of several maps of inbound and domestic tourist flows based on several years of statistical data. Mapping of the data enables us to look at the general patterns of data distribution, core and periphery patterns in choropleth maps, identification of clusters and outliers and possibilities of spatial autocorrelation. ArcGIS offers many options for the classification of data in the maps with different color schemes which can be helpful in highlighting different aspects of tourist data in the region. In this way, visualization of data in the form of maps provide base for further analysis of the data.

3.4 SPATIAL STATISTICS IN ARCGIS AND GEODA

The present research has employed the spatial statistics tools of ArcGIS program for the analysis of tourist patterns and trends in the selected region. Spatial analytic techniques have been showing continuous development since the advent of modern computer programs and associated software such as ArcGIS. Currently, GIS program is equipped with state of art tools of spatial statistics which are very useful in modelling and visualization of spatial data. Spatial statistics involves the concept of space (the area) in mathematical inferences which are absent in the traditional or nonspatial statistics. The statistical functions of ArcGIS provide useful analytic tools, categorized in four groups (Figure 3.2), to recognize patterns and clusters, geographic distribution and modelling spatial relationship of various variables (tourist patterns and trends in the present study).

Recent developments in the geographic information technologies have opened gateways for building innovative and user-friendly software equipped with spatial data analysis functions. GeoDa is one of the important spatial data analysis software which is very useful in visualization, mapping, and interpretation of spatial data. The added spatial statistics functions in GeoDa can be very practical in analyzing the large spatial

data in very little time. GeoDa is very similar to R in terms of functionality to spatial data analysis and powerful mapping techniques. It also provides the users point and click interface which is very easy to use. However, there are some limitations of GeoDa as it does not allow customization and modification of the input data according to the researcher's desire. Some analysis of the present study is made with the help of GeoDa software.
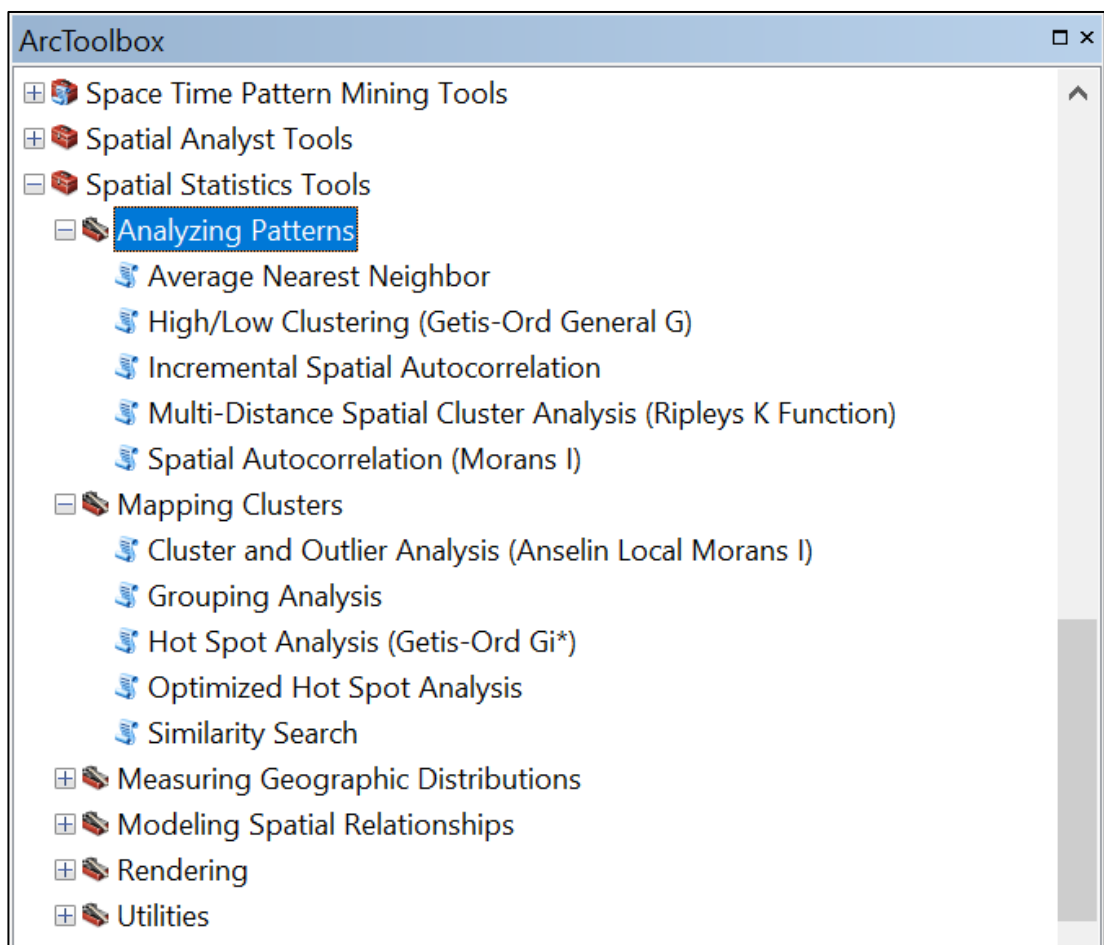

Figure 3.2: Spatial Statistic Tools in ArcGIS.

The spatial statistics tools of GIS have been used by many empirical studies of various fields to locate and explore the patterns of different spatial entities and phenomena. Several studies on tourism, discussed in the literature review in detail, have also tested

the method to acquire valid results regarding trends, patterns, and clusters of tourism activities in different regions of the world. In the current study, the spatial statistic tools of Analyzing Patterns and Mapping Clusters have been used to explore the extent of domestic and inbound tourists' activities in the region. Table 3.1 provides the detail of the functions which are applied in the study.

Table 3.1. Functions of different tools in ArcGIS spatial statistics.

Category	Tools	Function
Analyzing Patterns	High/low clustering (Getis-Ord general G)	Calculates the high and low values of tourist movements in the region
	Spatial autocorrelation (global Moran's I)	Finds the value of concentration or dispersion of tourists in terms of spatial autocorrelation through attribute data
Mapping Clusters	Cluster and outlier analysis (Anselin's local Moran's I)	Locate the presence of high or low clusters of tourists based on weighted features and also detect the spatial outliers
	Hot spot analysis (Getis-Ord i G*)	Identify the clusters of high and low tourist patterns of weighted set of features (tourist arrivals)

3.5 EXPLORATORY SPATIAL DATA ANALYSIS

After mapping the data and analyzing the choropleth maps of inbound and domestic tourism arrivals in the region, further analysis is made with the help of Exploratory Spatial Data Analysis techniques. The choropleth maps can only depict the core and periphery patterns of tourist flows in the region and no statistical inferences can be made. However, the maps can show some patterns of concentration and dispersion of tourist movements which require further analysis. ESDA has ability to find out the spatial clustering and outliers within the data. It has become an important analytic technique which is often used analysis based on spatial statistics, spatial econometrics, and geo-statistics. A very interesting point in the use of ESDA is the self-explanatory nature of used data instead of complex structuring before the analysis. Many empirical studies have tested and confirmed the usefulness of ESDA in understanding the patterns and trends of some selected variables over the space(Fu, Zhu, Kong, & Sun, 2012; Oxoli, Molinari, & Brovelli, 2018; Kang, Kim, & Nicholls, 2014; Haining, Wise, & Ma, 1998). Therefore, the present study has employed the ESDA to explore tourist movements in the region to highlight the past and current patterns and changes over time. ESDA techniques and functions have been added to various computer analysis software like R, Python, GeoDa, and ArcGIS. The present study has used GeoDa (Figure 3.3) and ArcGIS for the analysis of tourism data using ESDA techniques.
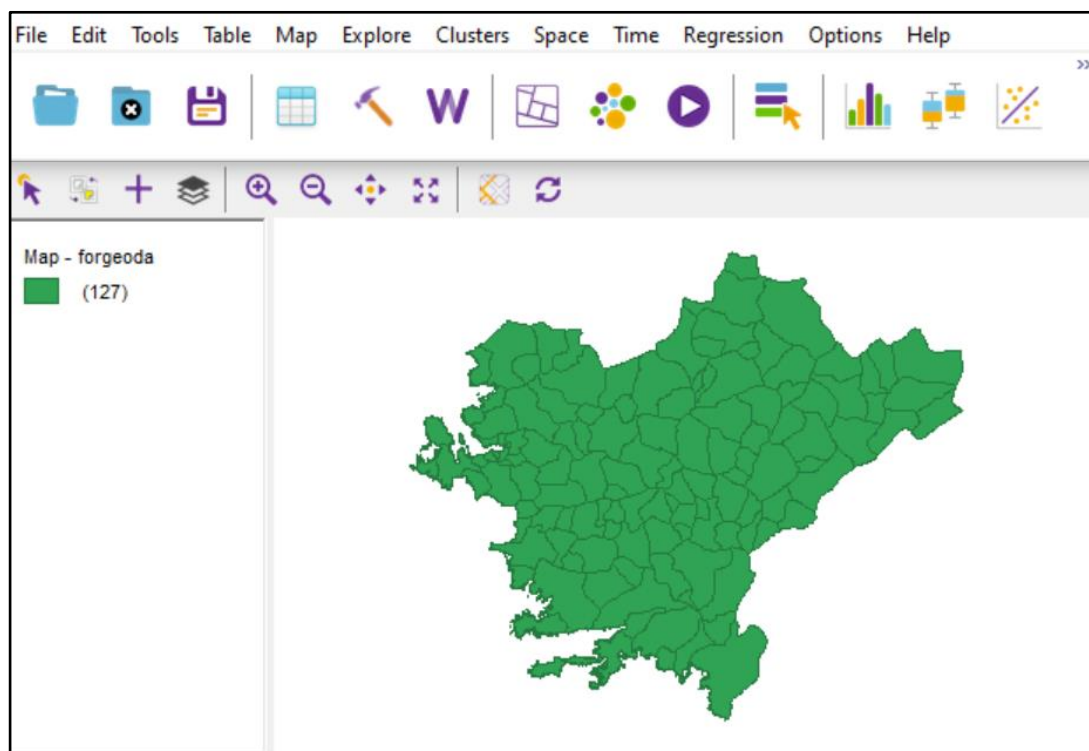
Figure 3.3: Map layout and functions in GeoDa

The quantitative exploration of tourism data in the current research is made by means of application of ESDA. The statistics tools provided by ArcGIS and GeoDA enable us to understand the spatial association of the tourist's movements in the Aegean region of Turkey. Several important analyses performed in the study are as follows:

- Spatial autocorrelation (global Moran's I)
- Cluster and outlier analysis (Anselin's local Moran's I)
- Hot spot analysis (Getis-Ord i G*)
- High/low clustering (Getis-Ord general G)

3.5.1 Spatial Autocorrelation

Spatial autocorrelation is a popular measure of describing geographical immediacy and association among different variables of interest found in the neighboring areas. In other words, the concept of spatial autocorrelation tells us about the possible connection of adjacent localities on map in which every spatial unit or component of area is considered as coded variable. The idea of spatial autocorrelation is thought to be stemmed from the Tobler's first law of Geography according to which, *'Everything*

is related to everything else, but near things are more related than distant things' (Tobler 1970, p.234). The concept of spatial autocorrelation is further explained by Hubert et al. (1981, p.224) as *'Given a set S containing n geographical units, spatial autocorrelation refers to the relationship between some variable observed in each of the n localities and a measure of geographical proximity defined for all n(n–1) pairs chosen from S.'*

The autocorrelation is measured in positive and negative values. It is assumed that if one representative matrix of all associations of any variable shows positive or negative relationship with another representative matrix of all associations of area components regardless of each other, there is positive or negative autocorrelation between them. Therefore, the two matrices would be similar in structure if both possess the association of neighboring high, low or zero values. If the values are clustered in a region that would indicate positive autocorrelation and if the values are dispersed the autocorrelation is negative, while 0 value depicts absence of any spatial association. The process is described in the Figure 3.4. A value of spatial autocorrelation which is greater than 0 and near to 1 shows strong positive autocorrelation. The ArcGIS program also helps visualizing and concluding the strength of autocorrelation.
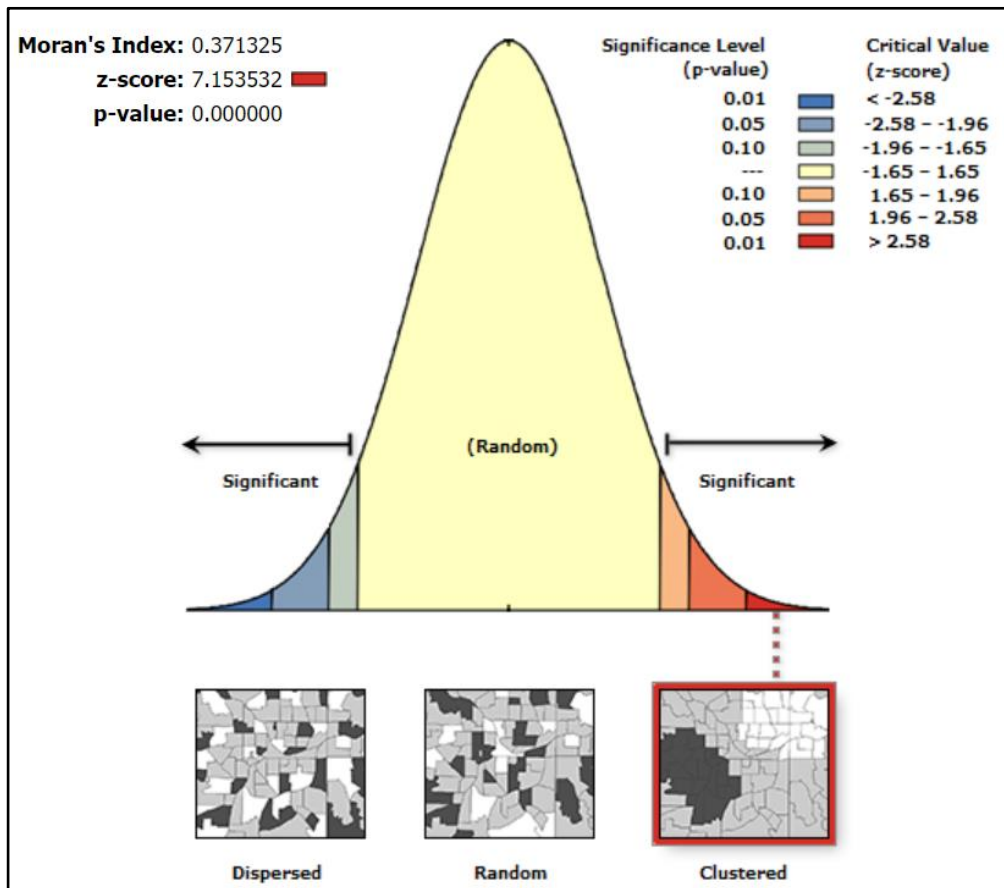
Figure 3.4. Description of Spatial Autocorrelation

3.5.1.1 Measurement of Spatial Autocorrelation - Global Moran's I Statistics

Spatial autocorrelation is difficult to calculate as it is very complex due to presence of multidimensional and bidirectional processes. Few statisticians have provided the mathematical foundations (also known as contiguity ratios) for the calculation of spatial autocorrelation among which Moran (1948), Krishna-Iyer (1949), and Geary (1954) are famous. Among the three measures of autocorrelation, Moran's I is favored and popularized due to its greater stability over other measures. In addition, the Moran's I measure is further analyzed and tested by Cliff and Ord (1981), which makes it more reliable measure of spatial autocorrelation. Therefore, the present study has also employed the Moran's I statistics in order to obtain the measures of spatial autocorrelation in tourist flows in the selected area.

The Moran's I statistics enables us to identify the presence of autocorrelation of tourism value of a spatial unit with itself and the spatial weight matrix, while no other variable is used. The Moran's I statistical measure is based on the weighted spatial matrices of district p and district q located in the selected study area. The measurement of Moran's I statistics is reached by identifying the attribute value and location of the variables (districts in the present study). The global measure of Moran's I statistics can be calculated by using the following formula (Cliff and Ord, 1981).

$$Moran\ I = \frac{N}{\beta_0} \frac{\sum p \sum q\ (w_{p,q})(Y_p Y_q)}{\sum p\ y_p^2} \tag{3.1}$$

Here, N denotes the total number of variables referring to the selected districts in current study. β_0 describes the sum of all the properties of selected variables. The symbol $w_{p,q}$ shows the spatial weights of a spatial unit (district in the present study) p and q, and the deviation of logged numbers of tourist arrivals (inbound and domestic) is represented by y_p. ArcGIS and GeoDa provide the preprogrammed functions of calculation of the Moran's I as shown in Figures 6 and 7. Spatial analysis is different from the traditional statistics analysis as it includes the parameters of spatial relationships in the calculations. Therefore, a value for the parameters of conceptualization of spatial relationships is required to select before performing the spatial analysis. Both GeoDa and ArcGIS offer different kinds of weight parameters to integrate in the analysis. The Inverse Distance parameter of spatial association with Euclidian distance method is selected for the data analysis in ArcGIS.

The Inverse Distance conceptualization is preferred in the analysis of clustering in any variable of interest. According to the conceptualization of Inverse Distance, all the selected variables or features affect all other features in a region (tourist arrivals in the present study) but the impact is greater on the nearby features or variables as compared to the farther ones. Similarly, GeoDa offers binary spatial weight matrices of rook and queen contiguity prior to the analysis of spatial association. The rook contiguity matrix calculates the nearby regions based on common borders only while queen contiguity also adds nodes in calculation of neighboring areas. In the present analysis Rook contiguity weight matrix is applied in the GeoDa analysis, although there are no significant differences in the results yielded by Rook and Queen contiguity.

3.5.2 Cluster and Outlier Analysis (Anselin's local Moran's I)

The presence of spatial autocorrelation in the tourist activities is determined by the use of global statistics of Moran's I, according to which, the data showed significant positive autocorrelation. However, the location of tourist clusters in the region cannot be identified with the use of global indicator of autocorrelation. Therefore, further analysis is necessary to identify the location of clusters and outliers in the tourism data. For this purpose, Local Indicator of Spatial Autocorrelation (LISA) is available in ArcGIS and GeoDa programs to visualize the high and low clusters of tourism activities on the maps which are also known as Moran's significance maps. Anselin, (2005) has provided the following formula to calculate the clusters and outliers in the data as Anselin!s Local Moran's I.

Local Indicator of Spatial Association ($LISA$) $I = \frac{p}{\sum p^2} \sum b \, w_{p,q}$ (3.2)

The value of spatial autocorrelation statistics ranges from -1 to +1 for determining the significance and strength of autocorrelation in the variables. A positive value of Moran's index which is close to +1 represents the significant positive correlation, while a negative value of the index closes to -1 demonstrates significant negative correlation. The 0 value of Moran's index means there is no autocorrelation in the data. In other words, a positive autocorrelation in the data indicates the clustered pattern and a negative autocorrelation represents the dispersed pattern while absence of autocorrelation indicates the random pattern in the data. We use the Mote Carlo permutation approach to measure the likelihood of equal observation in the data. Therefore, we conduct the re-estimation of LISA statistics for each permutation after shuffling the observed values in all locations randomly. In addition, during the analysis a reference distribution was also created by applying of 9999 random permutations to determine the significance level of LISA.

3.5.3 Hot spot analysis (Getis-Ord i G∗)

After measuring the global and local statistics of Moran's I, the study employed hotspot analysis to find out the important domestic and inbound tourist hotspots in the region for several selected years. The spatial statistics tool in ArcGIS is used to locate

the high and low clusters and hotspots of tourist flows. The significance of hotspots is based on the z-score (standard deviation) and p-value (probability value) which enables us to reject or accept the null hypothesis for the pattern analysis which is Complete Spatial Randomness (CSR), either of the features themselves or of the values associated with those features. In this context, a high value of z-score and small p-value for any feature in the analysis high clustering. In contrast, a small negative value of z-score and small p-value shows that low values are clustered. Therefore, higher (or lower) values of the z-score enables us to determine the intensity of the clustering. If the value of z-score is close to zero that means, there is no apparent spatial clustering.

The confidence level of the established hotspots is measured by the value of the field bin (Gi_bin) in ArcGIS software. The value of confidence level varies between -3 to +3. If the data set features fall in the +/-3 bins, it reflects higher statistical significance with a 99 percent confidence level. A 95% confidence level is confirmed when features fall in the values of +/-2. If the dataset features fall in the values of +/-1 bins, it reflects a 90 percent confidence level. The clustering for features in bin 0 is not statistically significant. The calculation of hotspot analysis is based on the following equation (Getis and Ord, 1992).

$$G_p^* = \frac{\sum_{q=1}^{n} w_{p,q} x_q - \beta \sum_{q=1}^{n} w_{p,q}}{S \sqrt{\left[\frac{n \sum_{q=1}^{n} w^2_{p,q} - \left(\sum_{q=1}^{n} w_{p,q} \right)^2}{n-1} \right]}} \tag{3.3}$$

In this equation x_q denotes the attribute value of feature q , (which indicates a district in the present study), $w_{p,q}$ is the spatial weight between district $p \ and \ q$, n is equal to the total number of districts. Whereas; $\beta \ and \ S$ are calculated as follows:

$$\beta = \frac{\sum_{q=1}^{n} x_q}{n} \tag{3.4}$$

$$S = \sqrt{\frac{\sum_{q=1}^{n} x_q^2}{n} - \beta^2} \tag{3.5}$$

3.6 PREPARING THE DATA FOR ARTIFICIAL NEURAL NETWORK MODELLING

The determination of domestic and inbound tourist hotspots is followed by the application of Artificial Neural Network techniques to predict the future locations for the tourism activities. In this study, a data driven feedforward neural network model is applied for further analysis of data.

Table 3.2. The selected factors for Feedforward Neural Network Analysis.

Classifications	Factors
Topography and Geomorphology	1. Geology 2. Aspect 3. Slope 4. Geological Faults 5. Coast Line 6. Altitude
Geomorphometry	7. Water Streams
Anthropogenic Activities	8. Land Use 9. Transportation

The data preparation stage of developing the NN model to produce projected hotspots of tourism activity involves producing several maps of factors apart from already produced hotspot maps. The hotspot maps produced by employing exploratory spatial data analysis technique is used as dependent variable in the model. Besides, some explanatory factors were collected from several sources such as topographic maps and digital elevation model (JAXA, 2019). The list of collected factors was classified and a list of all 9 factors are given in Table 3.2. These factors are selected after following

the literature on tourism studies (Romano et al. 2015; Deng et al. 2019; Kaptan Ayhan et al. 2020).

3.7 DEVELOPMENT OF FEEDFORWARD NEURAL NETWORK MODEL

The present study has developed feedforward neural network model which is also known as shallow neural network model. The feedforward neural network model is one of two main types of neural networks (Kim and Gofman, 2018; X.-D. Zhang, 2020). The second type is the deep neural networks. Shallow NN is only suitable for prediction of any spatial process (Land susceptibility, risk assessment or potential evaluation) in the future. The deep neural networks are more appropriate for identification of several spatial problems by detecting the location of any geographical phenomenon after their occurrence. However, the objective of the present research is to find the possibility of future tourism hotspots in the study area. Therefore, the use of feedforward neural network model is most suitable in the present study. Neural networks imitate the physical structure of the biological cell (dendrites, nucleus, and axons). In the artificial neuron, dendrites are represented by the input factors, Nucleus is represented by the artificial neurons and axons are represented by the connection between the neurons and forward or output layers.

Feed forward neural network model is developed by arranging a group of artificial neurons as input, hidden and output layers. In this way, the selected neural network model, as part of machine learning algorithm, is trained to map the future hotspots based on the former hotspots of tourism. The selected model of ANN can generate a weighted statistical model by extracting the geographical and geological parameters of each hotspot site, which can further map the possible areas for new hotspots. The main strengths of application of ANN model in tourism potential determination are based on its capability to handle several types of spatial data at different scales as well as ordinal and nominal data.
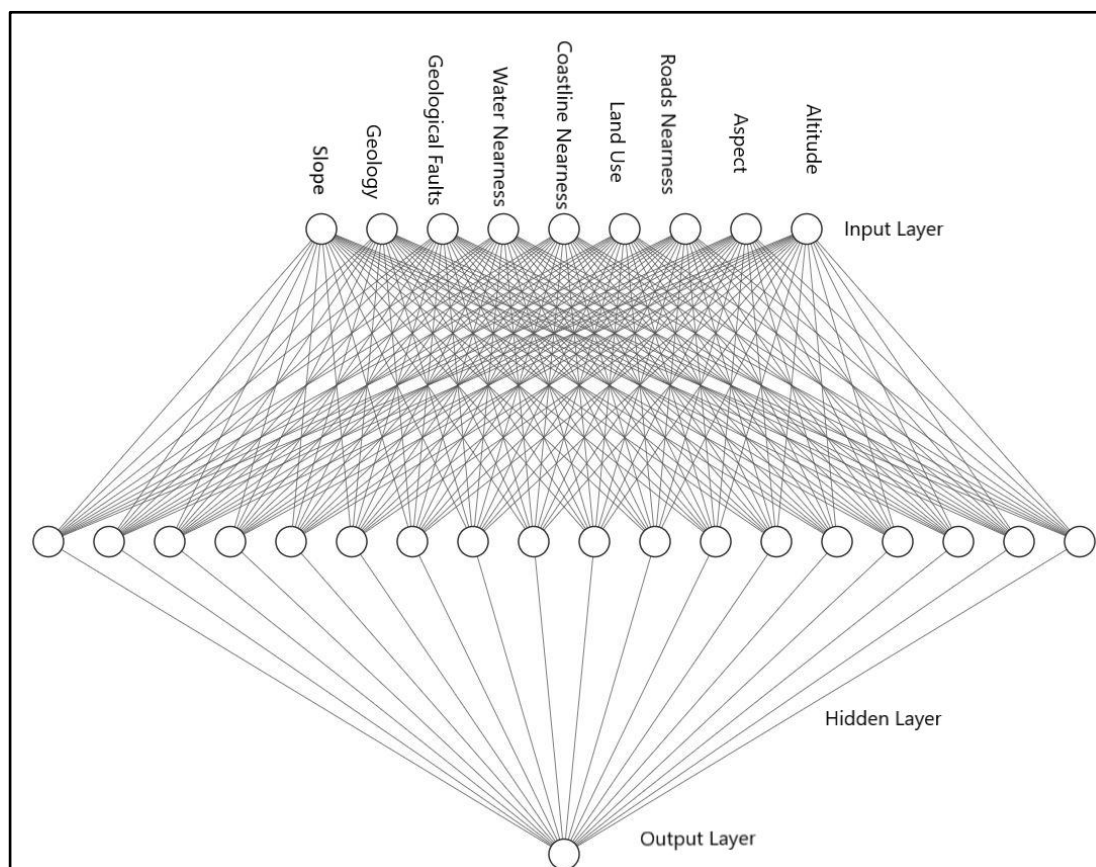
Figure 3.5. Development of Feedforward Neural Network Model.

The structure of feedforward NN model is displayed in Figure 3.5. The algorithm structure contains three layers. The first one is the input layer, which represents the input factors. The input factors are the relevant spatial data of the previous hotspots made through ESDA technique. The second layer is the hidden layer, which represents the layer of neurons. This is the most important layer where all operations of the analysis are performed. The hidden layer contains the artificial neurons that specify the ability of the model to extract the knowledge from the previous hotspot and factor dataset and build the model. The last layer represents the output layer. The output layer of the current algorithm is scaled between 0 and 1. The values represent the potential of future hotspots in the area. The developed model was generated in Matlab software.
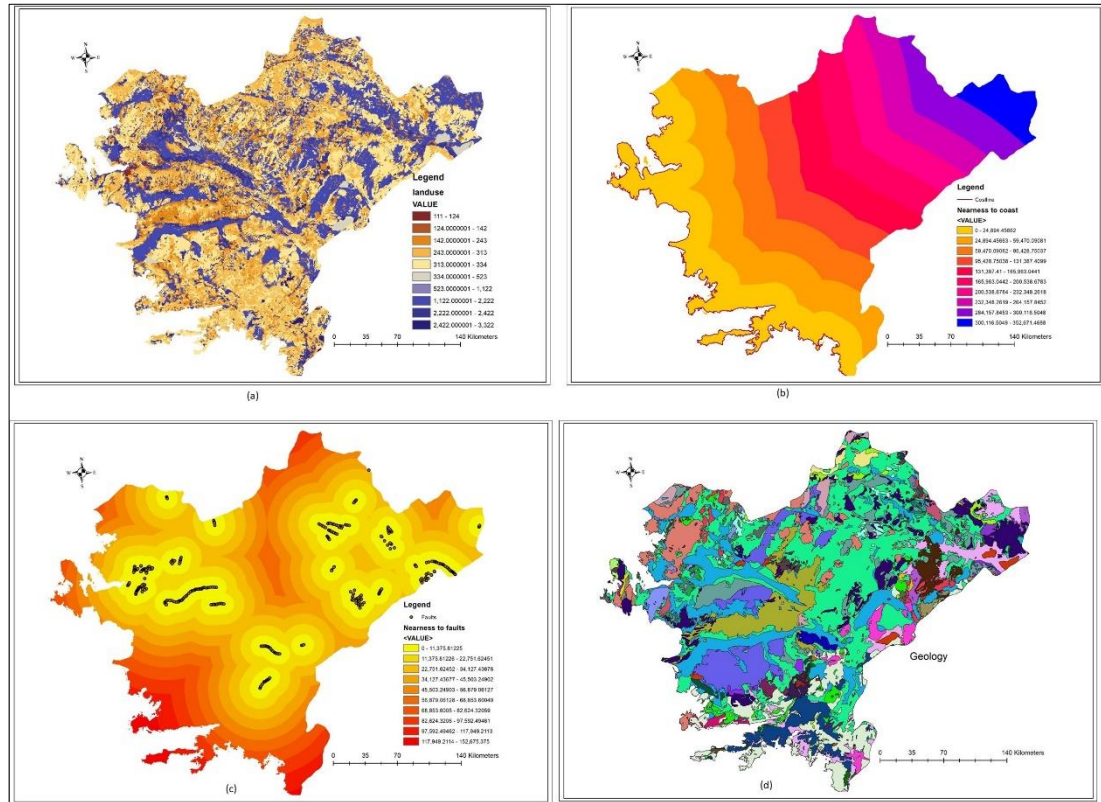
Figure 3.6. The factor maps of (a) land use, (b) nearness to coast, (c) nearness to fault lines, and (d) Geology.

The maps of selected factors for the analysis in NN model are presented in figures 3.6 and 3.7. Many previous studies have found land use, nearness to fault lines, nearness to coast and Geology of the region as important factors in determination of suitable tourism destinations. The tourism activity in the Aegean region is included in the mass tourism of 3S tourism i.e sand, sun, and sea. Therefore, coastal areas are potential hubs of tourist activities in that region. However, development of other destinations in the future requires to analyze the land use patterns of the region too. Meanwhile, new destinations should also consider the nearness to waterbodies like streams and water channels as they are considered as big attractions to tourists.

Geology of an area provides important information about the type of rocks and minerals of an area. Tourism related infrastructure building cannot be built without prior knowledge of the underlying characteristics of the bedrock and structures of the rocks. Planning for development of a stable tourist destination must also include the

41

information regarding Geology of the region to avoid problems related to natural disasters like earthquakes and tsunamis.
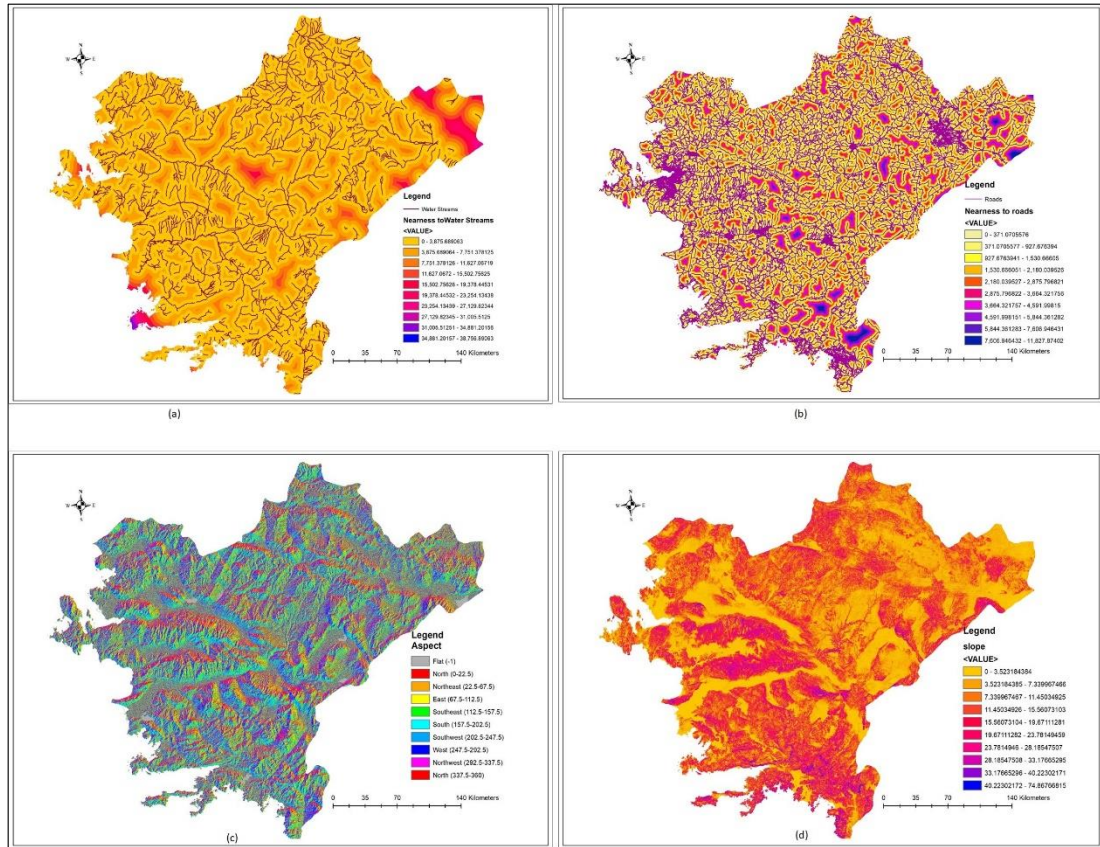

Figure 3.7. The factor maps of (a) nearness to water streams aspect, (b) nearness to roads (c) aspect, and (d) slope.

Availability of transport and nearness to roads is very crucial in planning of new tourist destination. Transportation is an important part of tourism services which can influence the tourists' choice of destination selection. Poor transport infrastructure may lead to negative image of the destination that can impact the economic growth related to tourism in the destination region. Therefore, any new tourism destination should be located to easily accessible areas through road networks. The factor of aspect is important in several climatic and physical impacts that can also indirectly affect tourism activity specially mountaineering and trekking.

CHAPTER 4

RESULTS AND DISCUSSION

The first part of analysis in the present study is conducted through ArcGIS and GeoDa software followed by the estimation of ANN model in MATLAB. In the mapping stage, choropleth maps of distribution of inbound and domestic tourist arrivals were made for realization of patterns of tourism activity in the region (Figure 4.1 and 4.2). In the analysis of choropleth maps, core and periphery areas of tourism distributions are observed. The core and periphery patterns demonstrate the presence of some core areas of tourism activities surrounding the low values of the tourist arrivals. There is considerable difference in the distributional patterns of domestic and inbound tourist arrivals. It is found that inbound core areas are mostly located in the coastal areas of the region, while most of the inland area falls in the periphery. On the other hand, domestic tourist arrivals show core and periphery patterns both in coastal and inland areas.

Simple choropleth maps can only help visualizing the distribution of tourism activities, without providing any information about spatial relationship and statistical inferences. These maps cannot depict the clusters of low and high tourist flows; neither can show the presence of spatial association and dependence of tourism in different districts of the region. Therefore, further analysis is made with the help of spatial statistics tools are to determine the presence of any spatial relationship within the data. According to the concept of spatial autocorrelation, the high values of any variable of interest in some area show spatial association to surrounding areas with high values. The results of spatial autocorrelation of tourism data over selected period reveal the presence of clustered patterns in both inbound and domestic tourist arrivals which is a sign of positive autocorrelation. Moran's I results of the present study confirmed the presence of significant positive autocorrelation in tourist flows analyzed in the selected region.
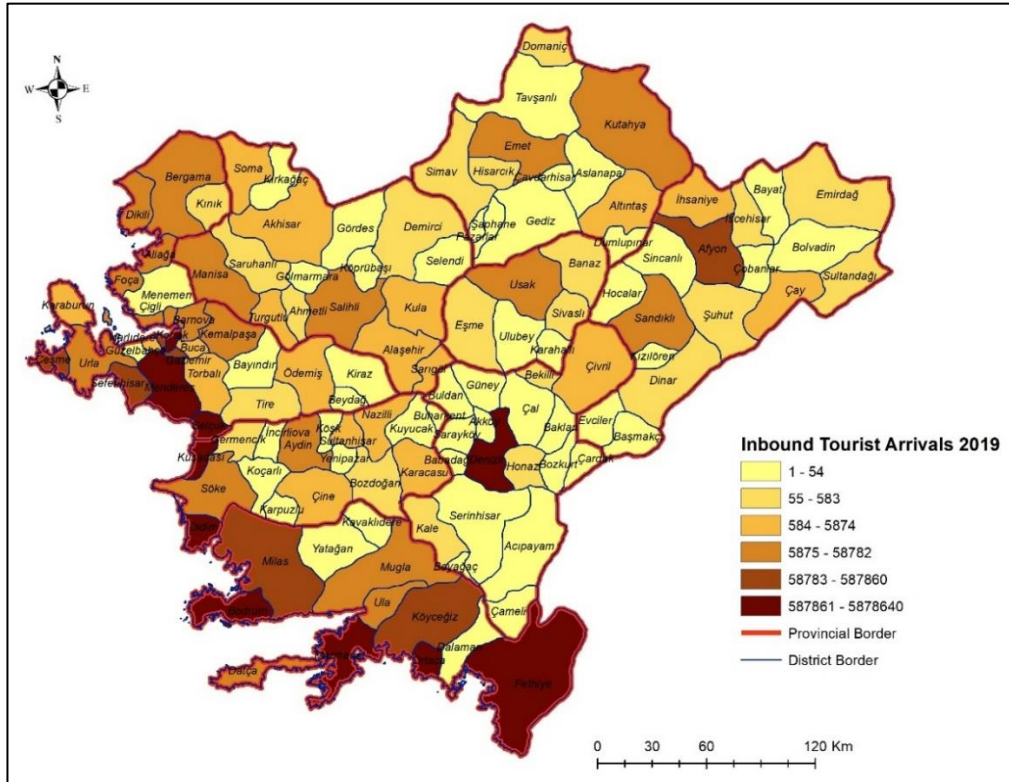
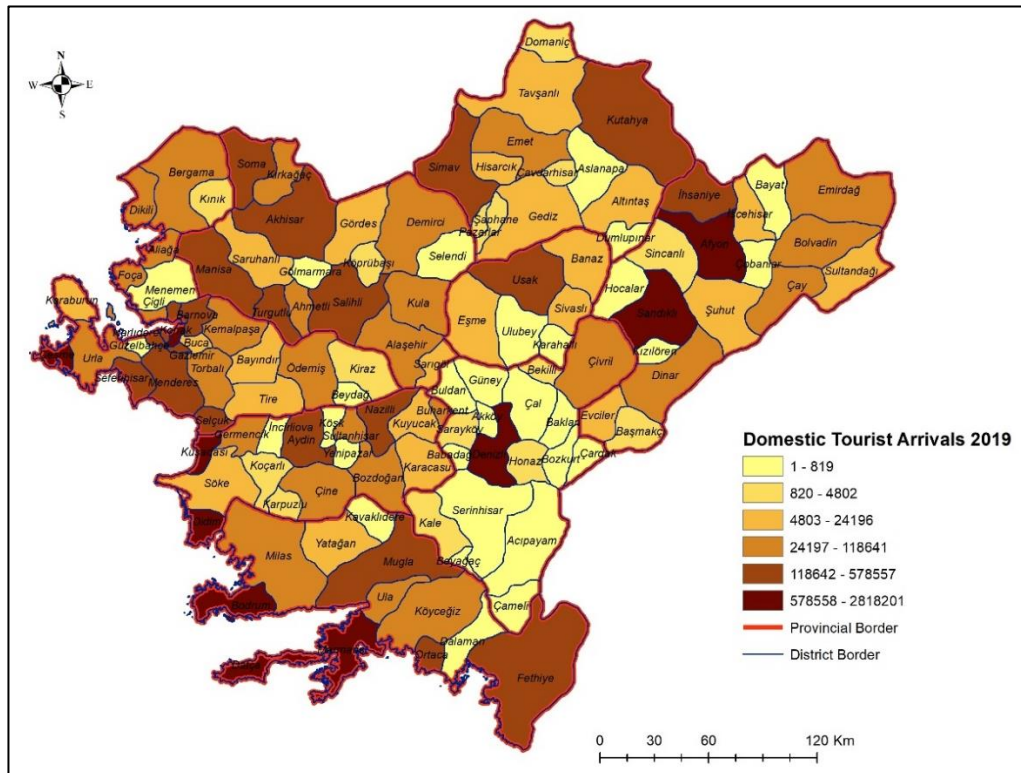Figure 4.1. The choropleth map of inbound tourist arrivals.

Figure 4.2 The choropleth maps of domestic tourist arrivals.

The value of Moran's I (Table 4.1) for inbound tourist arrivals is found more higher than domestic one. It means that inbound tourist arrivals show more strong and significant spatial association as compared to the domestic ones. In addition, the values of spatial autocorrelation statistics for different selected years also depict slight change in spatial association of tourist arrivals over time.

According to the results, strong positive values of Moran's I were observed in the years of 2015 and 2016 with little decrease of the autocorrelation in next years. The results have shown that there is a decline in the clustering of foreign tourist flows. The reason behind this change in inbound tourist patterns can be linked to emergence of new tourist destinations which play important role in dispersion of tourists. In contrast, a growing value of Moran's I in the domestic tourist arrivals have shown an increasing trend of spatial dependence with time. However, in general the domestic tourist arrivals show less clustering as compared to foreign ones. The literature has shown that the domestic tourism in Turkey is dispersed to other areas of the country including the Black Sea region. Therefore, the Moran's I values are not as strong as inbound tourists. However, the results of Moran I show an increasing tendency of domestic tourist clusters in the region.

The spatial autocorrelation test comprising global indicator of Moran's I statistics is only useful to detect the presence of spatial association and strength of the relationship between variables of interest. The global indicator is unable to show the exact locations of the clusters of tourist activities in the region. In order to identify the locations of the clusters, the LISA test was conducted which determined the location of the domestic and inbound tourist clusters. With the help of local statistics of Moran's I, significance maps were developed to visualize tourist clusters and hotspots of the domestic and inbound tourist arrivals in the region. The significance maps also helped in detection of the change in tourist clusters over time along with the visualization of the location of clusters. It is assumed that the change in the patterns of tourist clusters over time is related to the development of new tourist destinations and attractions which helped in dispersion of tourist activities in the region. It is already observed that Global Moran's I statistics showed very weak and positive autocorrelation in the domestic tourist

arrivals. The validation of global Moran's I results was done through significance maps of 2015 and 2019 in the visual form.

Table 4.1. The Moran's I values for domestic and inbound tourist arrivals.

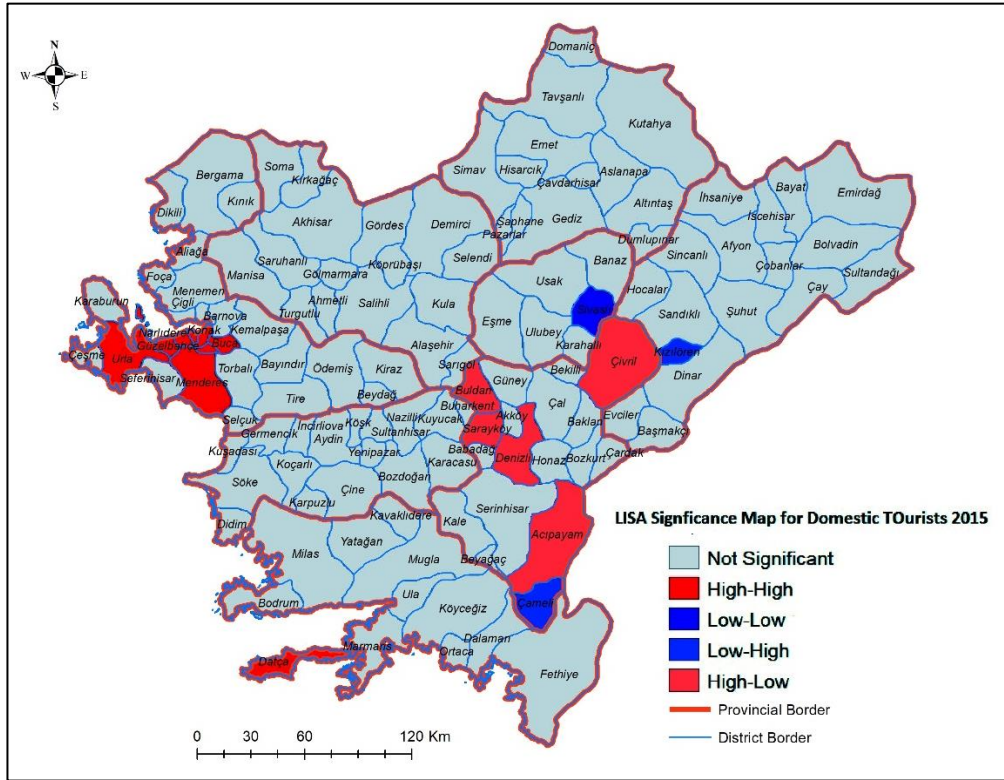| Years	Inbound Tourists		Domestic Tourist	
	Moran's I Value	P- Value	Moran's I Value	P-Value
2019	0.238	0.001	0.122	0.02
2018	0.240	0.001	0.11	0.02
2017	0.237	0.001	0.127	0.01
2016	0.308	0.001	0.171	0.003
2015	0.311	0.001	0.085	0.06

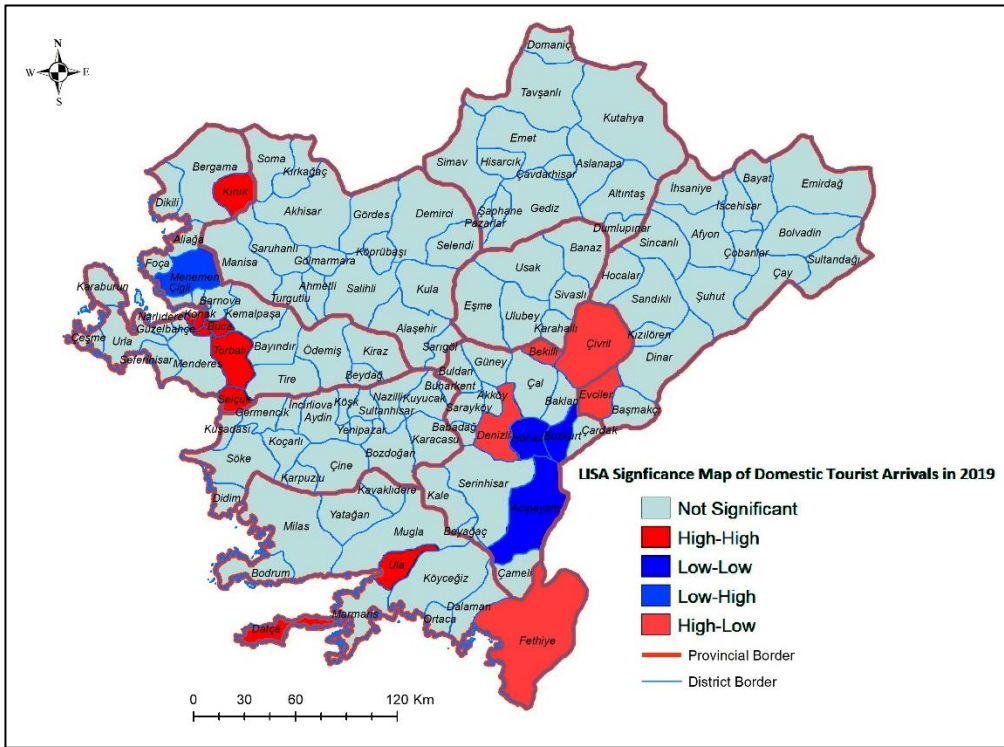Figure 4.3. The Moran's significance maps for domestic tourist arrivals 2015.

Figure 4.4. The Moran's significance maps for domestic tourist arrivals 2019.

The results of local indicator of Moran's I test in the form of significance maps have also shown gradual change in the domestic tourist clusters over time (Figure 4.3 and 4.4). It was found that the western parts of the region comprising the province of İzmir, faced little decline in the domestic tourist arrivals. In contrast, the tourist arrivals were found slight clustered in the province of Muğla. The results of LISA significance maps of inbound tourist arrivals have shown a strong correlation and high clustering of tourist activities in the region (Figure 4.5 and 4.6). The inbound tourist flows also showed change in the location of clusters over the selected years. The results of significance maps were in line to the results of global indicator of spatial autocorrelation showing slight decline in the values of Moran's I in recent years. It was found that the inbound tourist flows have identified clustering in the new districts of İzmir province. Additionally, the province of Aydın has also demonstrated the emergence of new inbound tourist clusters in the year 2019. Similarly, new clusters are also observed in the province of Muğla.
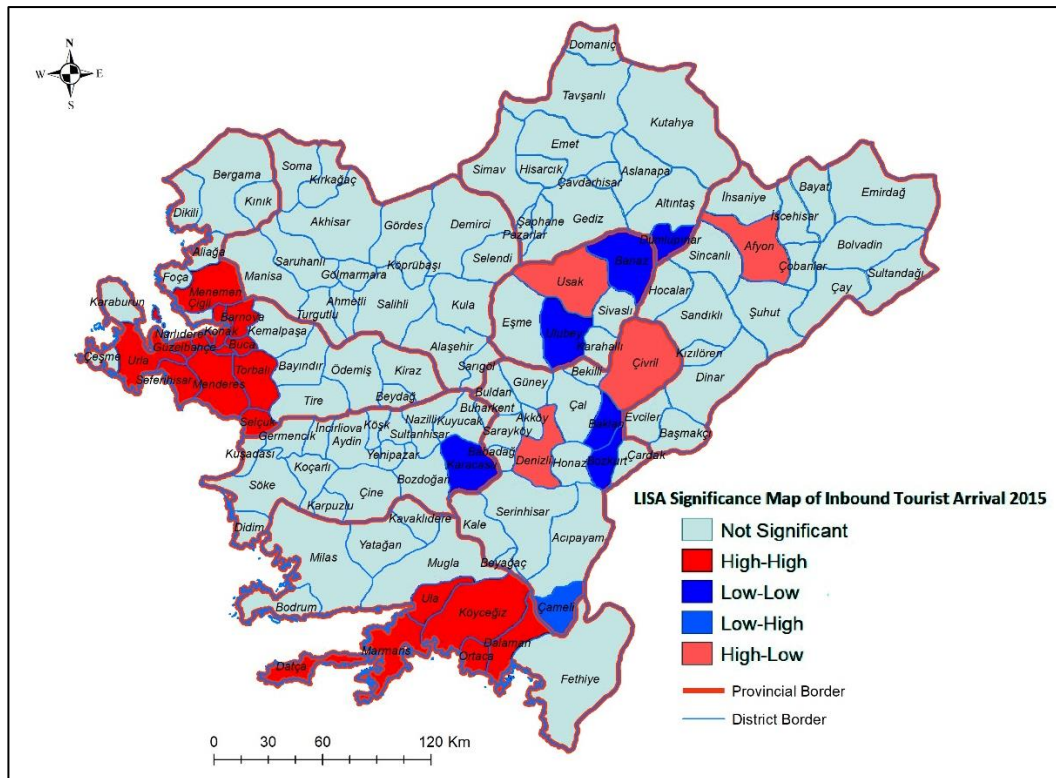

Figure 4.5. The Moran's significance maps for inbound tourist arrivals 2015.
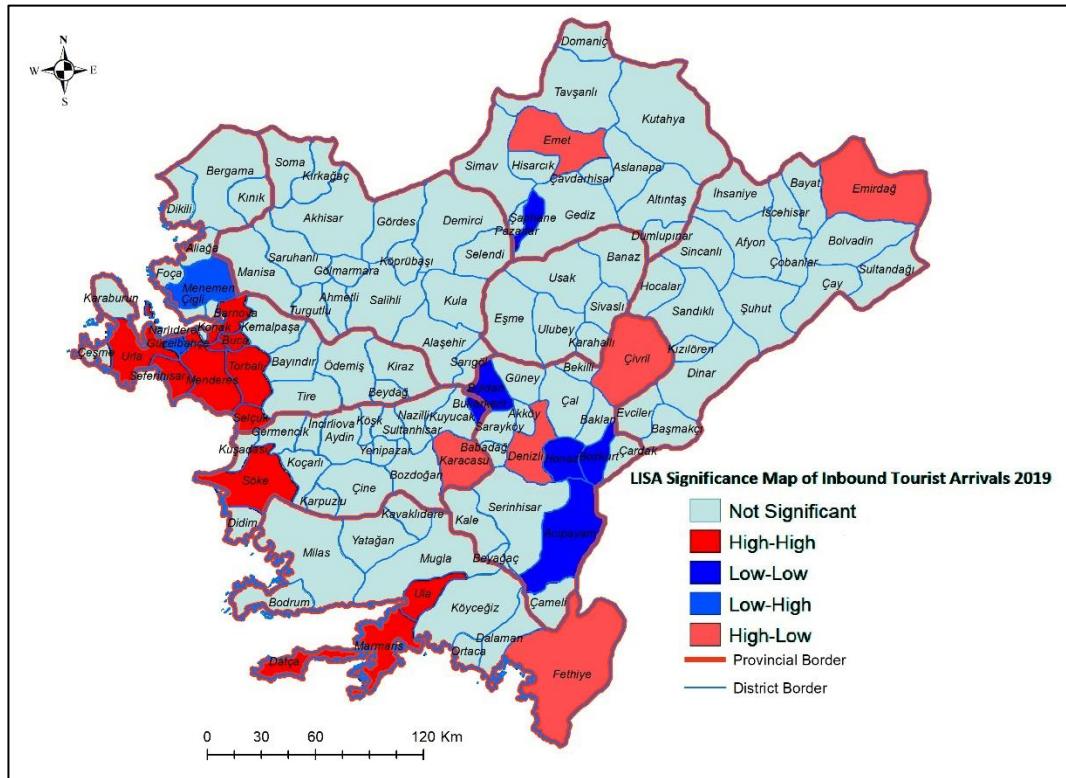
Figure 4.6. The Moran's significance maps for inbound tourist arrivals 2019.

Local indicator of spatial association (LISA) enables us to analyze the local clusters and outliers of tourist patterns in the region. The proximity of significant high and low values to other high and low values of tourist arrivals tells us about the possible areas of further tourism expansions. Further analysis of hotspots (Getis-Ord i G∗) reveals the locations of high tourist concentrations as hotspots and lower concentrations as cold spots (Figure 4.7).

According to hotspot analysis, the most concentrated areas in terms of domestic tourism activities in 2015 are located in western part of the Aegean Region including districts of province of İzmir. In contrast, the inland areas found in the eastern and north eastern parts of the region including districts in the province of Uşak, Afyonkarahisar and Denizli are considered as cold spots regarding domestic tourist flows. Although, the province of Denizli is famous for heritage tourism, the tourism activity is limited to some specific districts which did not show significant impact on the analysis. Some new areas located in provinces of Aydın and Muğla are also

observed as tourist hotspots in 2019, which show growth of domestic tourism activities in the region.
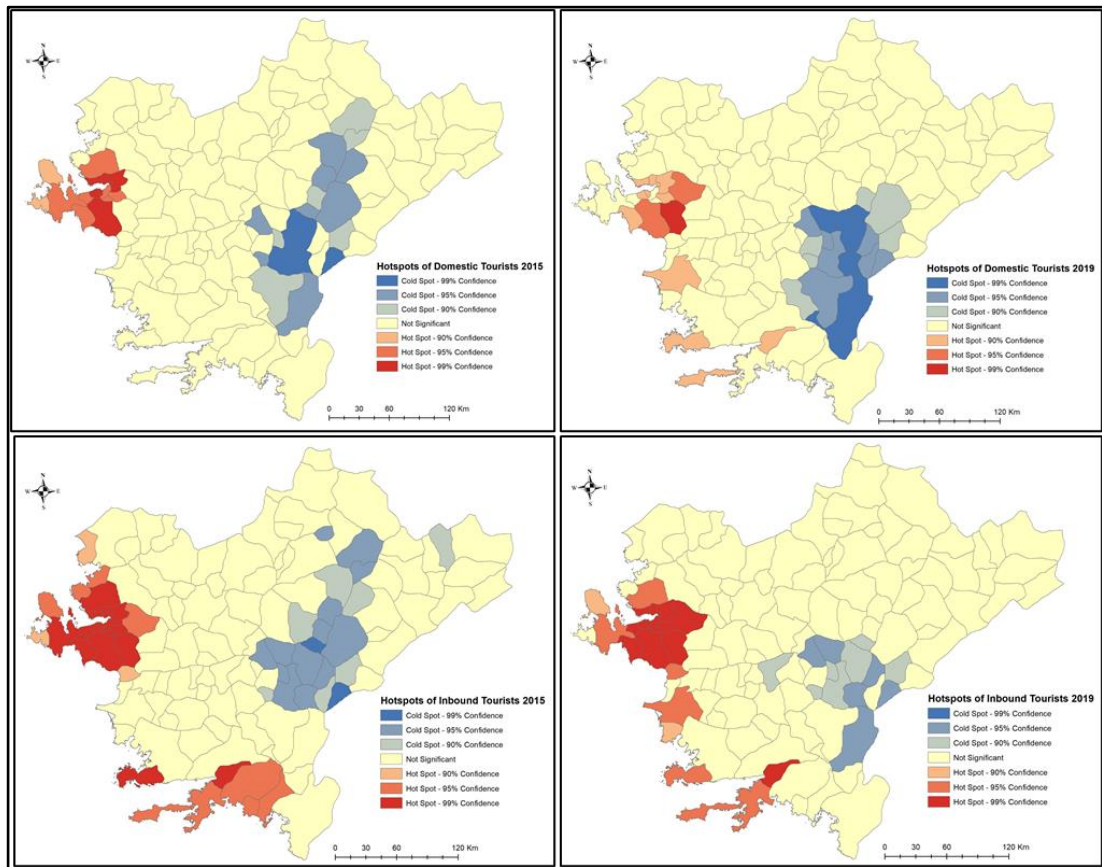

Figure 4.7. Hotspots of tourists (a) Domestic 2015 (b) Domestic 2019 (c) Inbound 2015 (d) Inbound 2019.

On the other hand, three important hotspots of inbound tourists are found in the region in 2015. The first hotspot is located in the coastal areas of province of İzmir, followed by the coastal areas of Aydın and Muğla. The cold spots are found in the inland areas just like domestic tourism. The hotspots of inbound tourism show a little growth in 2019 with increasing from three in 2015 to four hotspots in 2019. This conditions also depicts the growth of foreign tourists travelling to the region. The growth of tourism and changes in the tourist clusters is an important finding of the study which is helpful in exploring the important locations of tourist flows and changes in the patterns over time. It will further help in planning and management of tourist activities and resource allocation based on the priority areas.

After performing the spatial statistics analysis and getting the data related to tourism hotspots in the region, the ANN model was run for the prediction of potential tourist hot spots in the region. A correlation test was performed to validate the selected factors included in the ANN analysis. The results of Spearman's rho correlation tests are described in Tale 4.2. All the selected factors have found correlation with the tourist hotspots greater than 2% except the factor of slope. According to the literature, the factors having correlation greater than 2% is suitable to be considered in the analysis. Out of the eight factors, Geology, Nearness to Fault lines, Nearness to Coastline and Roads have shown more stronger correlation. Nearness to water shows a correlation value of 1.3% which reveals the insignificance of the presence of river or stream in occurrence of tourism hotspot. This is quite normal as the destination is located on the seaside.

Some of the factors show positive correlation with tourist hotspots, while others show a negative one. These positive and negative signs describe the potential of the existence of a tourist hotspot in relation to that specific factor. For example, nearness to coast shows negative correlation with the tourism hotspot which means farther location from the coast is not likely preferred as a tourist destination. Similarly, a destination far from the road is also not considered suitable for tourism activities due to importance of roads for transportation to and from the hotspot.

Table 4.2. The correlation between the tourism hotspots and the collected factors.

			Hotspots	Aspect	Slope	Roads	Water	Land-use	Geology	Faults	Coastline
Spearman's rho	Hotspots	Correlation Coefficient	1.000	.022	-.005	-.109[**]	-.013	-.023	.083[**]	.228[**]	-.293[**]
		Sig. (2-tailed)	.	.491	.882	.001	.678	.473	.008	.000	.000
		N	1000	1000	1000	1000	1000	1000	1000	1000	1000
**. Correlation is significant at the 0.01 level (2-tailed).											
*. Correlation is significant at the 0.05 level (2-tailed).											

All the selected factors provide the base of dataset to train the feedforward Neural Network model for prediction of future tourist hotspots in the region. The proposed neural network model is presented in Figure 4.8. During the process training the model, we use several training functions to train the NN model for the improvement of their performances. For this purpose, Levenberg-Marquardt backpropagation (Trainlm), Scaled Conjugate Gradient (Trainscg) and Bayesian regularization backpropagation (Trainbr) functions were applied. The best performance accuracy was obtained by the Trainlm function. In addition, further experiments were conducted to find the effects of the number of neurons on the performance accuracy. Based on the best training function (Trainlm), an optimizer developed to find the accuracy of the developed model using range of number of neurons in the hidden layer.
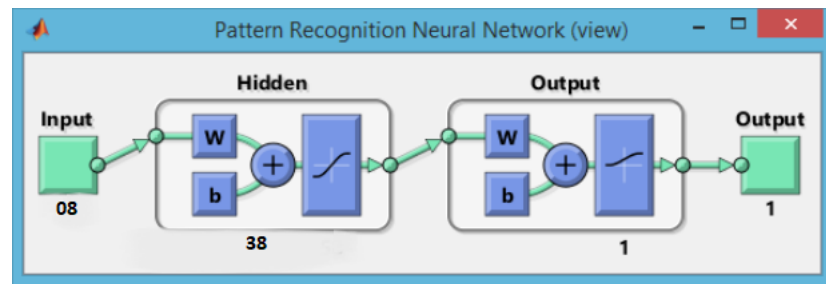

Figure 4.8. An example of the developed model structure based on all 08 factors.

In order to find the accuracy of the selected NN model was tested to achive best structure. Therefore, the developed NN model was checked using Confusion matrix metric for accuracy. The accuracy test was conducted by categorizing the dataset into training, validation, and testing datasets by use of 10-folds cross-validation method. Figure 4.9 illustrates the accuracy of selected model through the use of confusion matrix for training, validation, and testing of the dataset along with the overall model accuracy. It was revealed that the performance accuracy of the selected model based on the training sub dataset was 95.7 %. On the other hand, the performance accuracy on test confusion matrix was found 92.7%. The accuracy between training dataset and testing dataset showed that there was no over-fitting problem in the developed NN model since the testing accuracy was less than the training accuracy.

Figure 4.9. Confusion Matrix of NN model.

The final map of potential tourism hotspot is produced with the help of developed model of feedforward neural network Figure 4.10. All the maps of selected input factors were processed and extracted from ArcGIS 10.5 software to MATLAB R2019b as text files. In the MATLAB environment, the extracted text files were further analyzed with help of chunk processing in Python programming language. The study area was divided into 15 chunks. Meanwhile, predicted values of future tourism hotspots were obtained by processing the chunks datasets using the trained NN model. The obtained values were again exported to ArcGIS and reshaped to generate the final map of future tourism hotspots. The level of areas suitable for future tourism development ranges from 0 to 1.
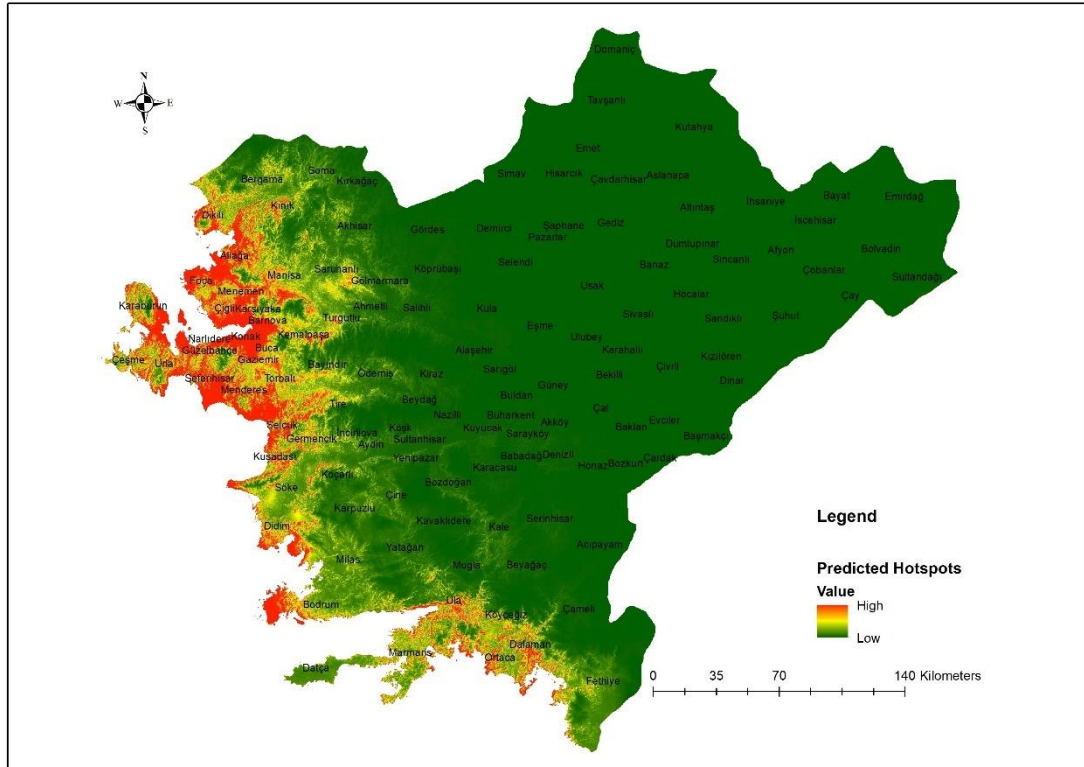
Figure 4.10. Map of suitable areas for future tourism development in Aegean Region of Turkey.

The resultant map of predicted tourism hotspots has provided large extent of the areas which are suitable for tourism activities in the region. Most of the future hotspot areas are found in the provinces of İzmir, followed by Muğla and Aydın. Among the districts of İzmir province, the most suitable areas of potential tourism hotspots are located in the districts of Konak, Foça, Aliağa, Bornova, Narlıdere, and coastal areas of Seferihisar, Menderes and Karaburun. In the Aydın province, Didim and Söke are found most suitable areas for future tourism development. Similarly, the coastal areas of Muğla province including Bodrum, Ula, Marmaris, Ortanca and Dalaman have great tourism development potential in the region.

A growing number of tourists in the region is seen as important factor of economic growth. However, such rapid tourism growth can also exert certain negative socioeconomic and environmental pressures if remained unplanned. Turkey is facing serious challenges in dealing the problems and issues related to rapid increase in the tourist arrivals in context of sustainable tourism development. It is argued that tourism related economic growth can face significant decline if no serious action is taken to

management of old tourist destinations and planning for the new ones. Butler (1980) has proposed a model of tourism lifecycle consisting of several stages of tourism development. According to the model, as the number of tourist arrivals increase the critical capacities of environmental elements, it needs serious management and planning to sustain the sectors, otherwise it will start to decline.

The present study has employed GIS and ANN techniques to analyze the tourism patterns and spatial dependence in the Aegean region of Turkey. There is large potential of using such techniques in further research on tourism development in other regions of Turkey. The study also suggests adding more factors including climate parameters, socioeconomic factors, and environmental properties for further analysis in future research.

CHAPTER 5

CONCLUSION

The application of GIS and ANN in research of various natural and social sciences has gained popularity in recent years. These computer techniques are proven very useful in managing and analyzing large scale spatial data of different natural, social, economic and environmental attributes. Recent growth in tourism and leisure activities over the globe due to availability of rapid transport and communication facilities has contributed significant share in global economy. Therefore, many developed and developing countries have shifted the attentions towards tourism development as important source of exchange earnings. New tourist destinations are being established to entertain more tourists for generating large revenues. However, the excessive utilization of tourism resources without any proper management and planning can lead to various socioeconomic and environmental problems. Therefore, a very cautious approach is required to develop new destination for maximizing the economic benefits of tourism without harming the surrounding natural environment. The analysis of spatial association and dependence of tourist flows is a prerequisite for better management and planning of tourism in any destination.

New developments in spatial data analysis techniques have provided great opportunities to map, visualize, and analyze the spatial data of tourism. The present study has employed the innovative techniques of GIS and ANN to analyze the patterns and spatial association of inbound and domestic tourist arrivals in the Aegean Region of Turkey. Turkey has become a popular destination of tourism in the world due to the country's unique geographical location, extensive coastal areas, favorable climate and other historical and cultural characteristics. The Aegean Region of Turkey is among the most famous regions of Turkey in terms of inbound and domestic tourism activities. the present research selected this region for determination of tourist patterns over the period 2015 to 2019. All the eight provinces comprising 123 districts are included in the spatial analysis to find the locations of high tourism concentrations.

The application of Exploratory Spatial Analysis (ESDA) techniques by using ArcGIS and GeoDa enables us to identify the spatial association of inbound and domestic tourist arrivals in the form of hotspots and cluster outliers. Results of global statistics of Moran's I have confirmed the presence of spatial autocorrelation in tourist arrivals in the region. The areas of high tourism activities are found clustered to other nearby areas of high tourist arrivals. The data is further analyzed by applying local statistics of Moran's I (also known as Anselin's local Moran's I) to identify the location of the tourist clusters. Several maps of significance were prepared to depict the current patterns and change in the inbound and domestic tourist clusters over time. It is revealed that the western parts of the region, including the province of Izmir, have shown a decline in the clustering of domestic tourist activities over time. In contrast, the significance maps of inbound tourist arrivals also show a strong correlation and high clusters of tourist activities in the region. Besides, a change in the clusters of inbound tourists is also observed over time. That might explain the reason for the slight decline in the values of Moran's I in recent years.

Further analysis of the tourism data by applying hotspot analysis technique revealed that domestic tourism activities in 2015 were concentrated in western part of the Aegean Region including districts of province of Izmir. On the other hand, the inland areas found in the eastern and north eastern parts of the region including districts in the province of Uşak, Afyonkarahisar and Denizli were found as cold spots. Hotspot analysis also locate some new areas in provinces of Aydın and Muğla as tourist hotspots in 2019, which show growth of domestic tourism activities in the region. Similarly, hotspots of foreign tourists were also found in the coastal areas of province of İzmir, followed by the coastal areas of Aydın and Muğla. The hotspots of inbound tourism show a little growth in 2019 comparing to previous years. The results of the current study are considered helpful in exploring the important locations of tourist flows and changes in the patterns over time. It will further help in planning and management of tourist activities and resource allocation based on the priority areas.

Based on the results of spatial autocorrelation analysis, a map with predicted future tourism hotspots was produced employing feedforward neural network model which provides useful information regarding location of best suitable destination. The

57

explanatory factors including land use, nearness to coast, nearness to fault lines, Geology, aspect, nearness to water streams, nearness to roads, and slope were selected to train the model. The final map describes the extent of possible future hotspots in the region. Most of the regions are found in the coastal areas of provinces of İzmir, Muğla and Aydın.

Tourism destinations of the coastal areas of Aegean region of Turkey are facing serious problem of overcrowding which leads to many socioeconomic and environmental problems. Therefore, a strong tourism development policy is required for the sustainable development of tourism focusing on minimum environmental damages and maximum economic benefits. In order to develop a robust tourism policy, the spatial extent and patterns of current tourism activities are necessary to evaluate. Modern day computer techniques with sophisticated software and innovative programs provide great opportunity to analyze big tourism data and its association to spatial patterns of tourist destinations. Hence, the findings of present study would be very useful for policy implications in the future tourism development in the Aegean region of Turkey.

The results of the study provide insights into the tourist flows and changes in the spatial association of tourist clusters. It is expected that the findings of the present study will provide useful information for future resource allocation in the region regarding tourism development.

REFERENCES

1. Aggarwal, Charu C. "Neural Networks and Deep Learning", *IBM T. J. Watson Research Center International Business Machines Yorktown Heights*, NY, USA, 1-20 (2018).

2. Anselin L "Local indicators of spatial association – LISA". **Geography Anal** 27(2):93-115 (1995).

3. Blazquez, C. A., & Celis, M. S. "A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile". *Accident Analysis and Prevention*, *50*, 304–311(2013).

4. Brida, J. G., Cortes-Jimenez, I., & Pulina, M. "Has the tourism-led growth hypothesis been validated? A literature review". *Current Issues in Tourism*, *19*(5), 394–430 (2016).

5. Cang, S. "A non-linear tourism demand forecast combination model". *Tourism Economics*, *17*(1), 5–20 (2011).

6. Chen, C. F., Lai, M. C., & Yeh, C. C. "Forecasting tourism demand based on empirical mode decomposition and neural network". *Knowledge-Based Systems*, *26*, 281–287 (2012).

7. Chhetri, A., Arrowsmith, C., Chhetri, P., & Corcoran, J. "Mapping Spatial Tourism and Hospitality Employment Clusters: An Application of Spatial Autocorrelation". *Tourism Analysis*, *18*(5), 559–573(2013).

8. Chhetri, Anjali, Chhetri, P., Arrowsmith, C., & Corcoran, J. "Modelling tourism and hospitality employment clusters: a spatial econometric approach". *Tourism Geographies*, *19*(3), 398–424(2017).

9. Claveria, O., & Torra, S. "Forecasting tourism demand to Catalonia: Neural networks vs. time series models". *Economic Modelling*, *36*, 220–228 (2014).

10. Cliff AD, Ord JK "Spatial processes: models and applications". *Pion, London* (1981)

11. Çuhadar, M., Iclal, C., & Ceyda, K. "Modelling and Forecasting Cruise Tourism Demand to İzmir by Different Artificial Neural Network Architectures". *International Journal of Business and Social Research*, *4*(3), 12–28 (2014).

12. Deng, J., Che, T., Xiao, C., Wang, S., Dai, L., & Meerzhan, A. "Suitability analysis of ski areas in China: An integrated study based on natural and socioeconomic conditions". *Cryosphere*, *13*(8), 2149–2167 (2019).

13. Eftekhari, E., & Mahdavi, M. "Land suitability assessment using ANP in a GIS environment for Tourism Development Site (Case study: Lavasan-e Kuchak Rural District, Tehran province, Iran)" *Journal of Tourism & Hospitality Research* . *7*(1), 5–17(2019).

14. Fernandes, P. O., Teixeira, J. P., Ferreira, J., & Azevedo, S. "Training neural networks by resilient backpropagation algorithm for tourism forecasting". *Advances in Intelligent Systems and Computing*, *220*, 41–49 (2013).

15. Fu, Q., Zhu, Y., Kong, Y., & Sun, J. "Spatial analysis and districting of the livestock and poultry breeding in China". *Journal of Geographical Sciences*, *22*(6), 1079–1100 (2012).

16. Getis A, Ord JK "The analysis of spatial association by distance statistics". **Geogr Anal** 24(3):189-206 (1992)

17. Gomes, M. J. T. L., Cunto, F., & Silva, A. R. "Geographically weighted negative binomial regression applied to zonal level safety performance models". *Accident Analysis and Prevention*, *106*(February), 254–261(2017).

18. Göymen, K. "Tourism and governance in Turkey". *Annals of Tourism Research*, *27*(4), 1025–1048 (2000).

19. Gregorić, M., & Baldigara, T. "Artificial neural networks in modelling seasonal tourism demand - case study of Croatia". *Zbornik Veleučilišta u Rijeci*, *8*(1), 19–39 (2020).

20. Haining, R., Wise, S., & Ma, J. "Exploratory spatial data analysis in a geographic information system environment". *The Statistician*, *47*(3), 457–469(1998).

21. Hamzaçebi, C. "Improving artificial neural networks' performance in seasonal time series forecasting". *Information Sciences*, *178*(23), 4550–4559(2008).

22. Hsu, L. C., & Wang, C. H. "Applied multivariate forecasting model to tourism industry". *Tourism*, *56*(2), 159–172(2008).

23. Hu, X., Neelam, C., & Green, G. "Spatial analysis of international tourism growth in China's cities". *Current Politics and Economics of Northern and Western Asia*, *25*(2), 149–175 (2016).

24. Hubert LJ, Golledge RG, Costanzo CM "Generalized procedures for evaluating spatial autocorrelation". *Geogr Anal* 13(3):224-233(1981).

25. Huarng, K. H., Yu, T. H. K., & Parellada, F. S. "An innovative regime switching model to forecast Taiwan tourism demand". *Service Industries Journal*, *31*(10), 1603–1612(2011).

26. Kang, S., Kim, J., & Nicholls, S. "National Tourism Policy and Spatial Patterns of Domestic Tourism in South Korea". *Journal of Travel Research*, *53*(6), 791–804 (2014).

27. Kaptan Ayhan, Ç., Cengiz Taşlı, T., Özkök, F., & Tatlı, H. "Land use suitability analysis of rural tourism activities: Yenice, Turkey". *Tourism Management*, *76*(February 2018) (2020).

28. Khan, A. A. "The spatial distribution and relationship of tourist flow in Turkey". *European Journal of Tourism Research*, *19*, 40–55 (2018).

29. Li, Y., & Cao, H. "Prediction for Tourism Flow based on LSTM Neural Network". *Procedia Computer Science*, *129*, 277–283(2018).

30. Malik, Nitin. "Artificial Neural Networks And Their Applications" National Conference on 'Unearthing Technological Developments & their Transfer for Serving Masses' *GLA ITM, Mathura, India* 17-18 April (2005).

31. Martins, L. F., Gan, Y., & Ferreira-Lopes, A. "An empirical analysis of the influence of macroeconomic determinants on World tourism demand". *Tourism Management*, *61*, 248–260(2017).

32. Moutinho, L., Huarng, K. H., Yu, T. H. K., & Chen, C. Y. "Modeling and forecasting tourism demand: The case of flows from Mainland China to Taiwan". *Service Business*, *2*(3), 219–232 (2008).

33. Nawaz, Z., Li, X., Chen, Y., Guo, Y., Wang, X., & Nawaz, N. "Temporal and spatial characteristics of precipitation and temperature in Punjab, Pakistan". *Water (Switzerland)*, *11*(9), 1–23(2019).

34. Oxoli, D., Molinari, M. E., & Brovelli, M. A. "Hotspot analysis, an open source GIS tool for exploratory spatial data analysis: Application to the study of soil consumption in Italy". *Rendiconti Online Societa Geologica Italiana*, *46*(February 2020), 82–87 (2018).

35. Pan, X., Yang, Z., Han, F., Lu, Y., & Liu, Q. "Evaluating potential areas for mountain wellness tourism: A case study of Ili, Xinjiang province". *Sustainability (Switzerland)*, *11*(20) (2019).

36. Romano, G., Dal Sasso, P., Trisorio Liuzzi, G., & Gentile, F. "Multi-criteria decision analysis for land suitability mapping in a rural area of Southern Italy". *Land Use Policy*, *48*, 131–143(2015).

37. Saeidi, S., Mohammadzadeh, M., Salmanmahiny, A., & Mirkarimi, S. H.

"Performance evaluation of multiple methods for landscape aesthetic suitability mapping: A comparative study between Multi-Criteria Evaluation, Logistic Regression and Multi-Layer Perceptron neural network". *Land Use Policy*, *67*(May), 1–12 (2017).

38. Stankov, U., Armenski, T., Klauco, M., Pavluković, V., Cimbaljević, M., & Drakulić-Kovačević, N. "Spatial autocorrelation analysis of tourist arrivals using municipal data: A Serbian example". *Geographica Pannonica*, *21*(2), 106–114 (2017).

39. Talebi, M., Majnounian, B., Makhdoum, M., Abdi, E., Omid, M., Marchi, E., & Laschi, A. "A GIS-MCDM-based road network planning for tourism development and management in Arasbaran forest, Iran". *Environmental Monitoring and Assessment*, *191*(11) (2019).

40. Tobler WR "A computer movie simulating urban growth in the Detroit region". **Econ Geogr** 46(2):234-240(1970).

41. Tosun, C. "Challenges of sustainable tourism development in the developing world: The case of Turkey". *Tourism Management*, *22*(3), 289–303(2001).

42. Wiley, J. "The Formation of the Tourist' s Loyalty How Does It Affect Price Discounts?" *International Journal of Tourism Research*, *9*(August), 453–464 (2007).

43. Xu, D., Huang, Z., Hou, G., & Zhang, C. "The spatial spillover effects of haze pollution on inbound tourism: evidence from mid-eastern China". *Tourism Geographies*, *22*(1), 83–104 (2020).

44. Yang, H., Dong, X., Wang, M., & Guo, Y. "GIS-Aided Evolvement Analysis of Spatial-temporal Pattern of Regional Tourism Industry Environment". *Advanced Materials Research 731*, 4690–4693(2013).

45. Yang, Y., & Wong, K. K. F. "Spatial Distribution of Tourist Flows to China's Cities". *Tourism Geographies*, *15*(2), 338–363(2013).

RESUME

Asma RAFIQUE was born in Islamabad Pakistan in 1982. She completed her school education in Islamabad. After high school she got admission in WISE College for graduation in computer sciences and completed her bachelor's degree from Allama Iqbal Open University Islamabad. She did her internship in Database and Communication in ACTIONAID in 2007. She worked as Mathematics teacher in Nisar Model School Rawalpindi from 2009 to 2011. Afterwards she moved to Ankara and worked in Pakistan Embassy International Study Group, Ankara, Turkey as ICT teacher from 2014 to 2016. She moved to Karabük and successfully completed Masters in Computer Engineering in 2021.

CONTACT INFORMATION

Address: Institute of Graduate Programs, The Department of Computer Engineering Demir-Çelik Campus/KARABUK

Email: asmamscs@gmail.com