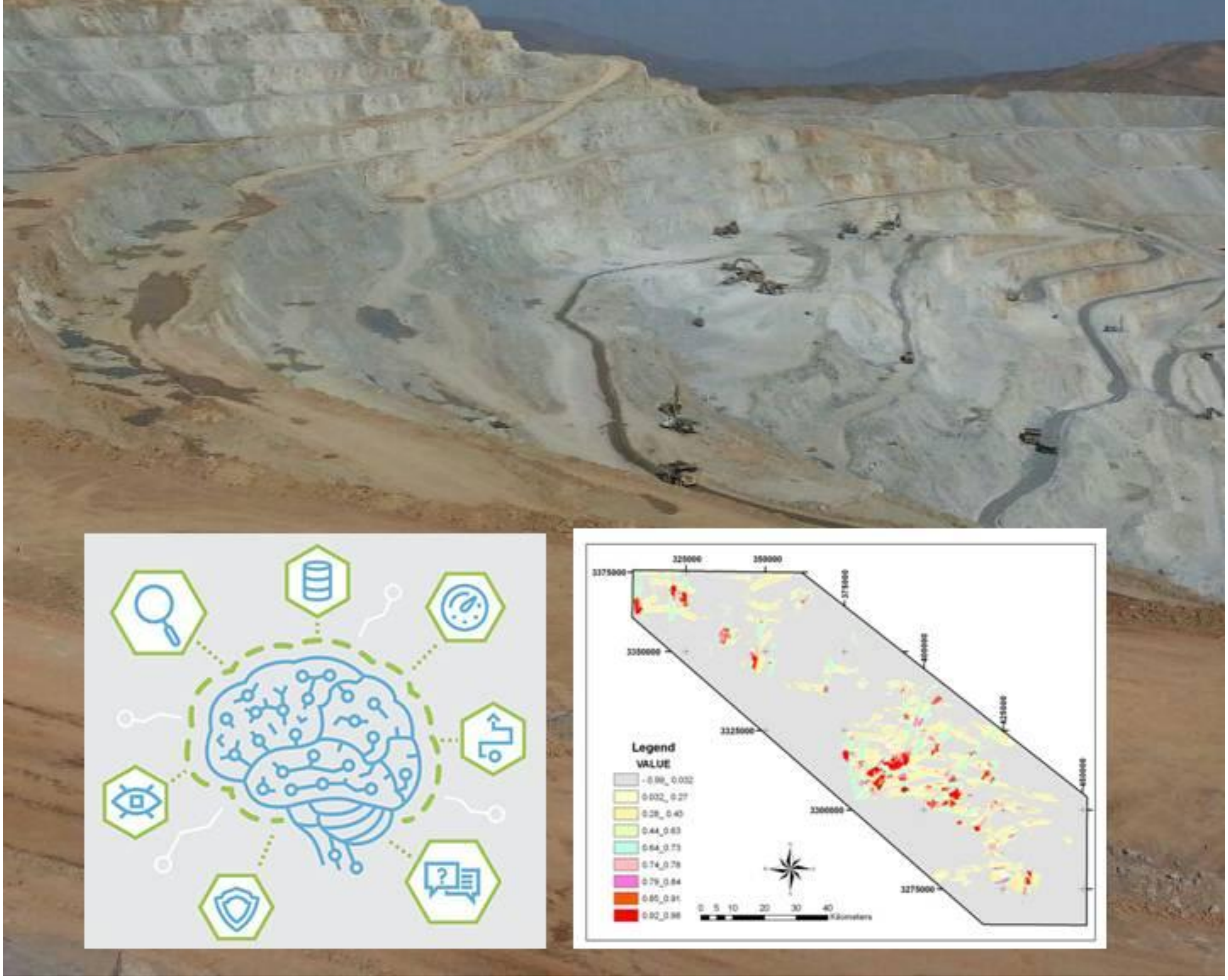
Machine Learning for Mineral Prospectivity Mapping

Alireza Jafari Rad

Case study: Porphyry Copper MPM Using neural networks in central Iran



Abstract

Machine learning is a data-driven algorithms and techniques that automate clustering, classification and prediction of data. Machine learning algorithms use computational methods to learn information directly from data without relying on a predetermined equation as a model. The algorithms adaptively improve their performance as the number of samples available for learning increases.

Machine learning algorithms find natural patterns in data that generate insight and help you make better decisions and predictions. With the rise in big data, machine learning has become particularly important for solving problems in areas.

Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data.

The aim of supervised machine learning is to build a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data supervised learning uses classification and regression techniques to develop predictive models.

Unsupervised learning finds hidden patterns or intrinsic structures in data. It is used to draw inferences from datasets consisting of input data without labelled responses. Clustering is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data.Choosing the right algorithm can seem overwhelming there are dozens of supervised and unsupervised machine learning algorithms, and each takes a different approach to learning. There is no best method or one size fits all. Finding the right algorithm is partly just trial and error even highly experienced data scientists can’t tell whether an algorithm will work without trying it out. But algorithm selection also depends on the size and type of data you’re working with, the insights you want to get from the data, and how those insights will be used (© 2016 The MathWorks, Inc.).

Consider using machine learning when you have a complex task or problem involving a large amount of data and lots of variables, but no existing formula or equation.

Neural Network inspired by the human brain. A neural network consists of highly connected networks of neurons that relate the inputs to the desired outputs. The network is trained by iteratively modifying the strengths of the connections so that given inputs map to the correct response (© 2016 The MathWorks, Inc.).

Artificial Neural Network (ANN) was employed in this research for Mineral Prosoectivity Mapping (MPM). ANN is a data-driven method, that the allocation of appropriate weights to data layers implement basis on the datasets. A Multi-Layer perceptron (MLP) neural network is applied for this research because it performs very well for a large variety of problem types and has powerful function approximation capabilities. Unlike other data-driven methods which need only one set of training points i.e. the site of mineral deposits, ANN requires another set of training points which defines the absence of mineral deposits. The second set of training points was generated using simple overlaying method. Also the operations for the application of neural networks employed in this study involve multi-class representation of evidential maps. The predictive map is generated from multi-class evidential maps show higher predictive rate with respect to known deposits and indications compared to the predictive map resulted from binary maps (Jafari Rad, 2009). This method is based on the applying GIS techniques to preparation of data layers, and also modeling using MATLAB software.

The study area of this research is the most important part of Cu (Mo) porphyry - type mineralization belt, where is located in central Iran. There are some well – known porphyry copper deposits in this region like Sar Cheshmeh and Meiduk mines, but certainly there are same grounds to search for new porphyry deposits.

CONTENTS

1- Introduction

1-1- Background to the research

Iran is located in Alpine-Himalaya orogenic and metalogenic belt formed after Tethys collision, and therefore has a high potential for different types of minerals. Although conventionally a unique Volcano – Plutonic – Arc (VPA) is considered to be formed by subduction of Mesozoic Tethys oceanic crust, but new evidences show that there are different oceanic basins, and associated arcs. One of the most important VPA is Kalkafi – Sar Cheshmeh– Kharestan (Samani & Ashtari 1992), which is located parallel to the Makran – Baft – Naeen ophiolitic belt, on the margin of central Iranian median mass and the study area of this research (Chahfiruzeh – Kuh Panj district).

This VPA is the most important Cu (Mo) porphyry - type ore belt extended from Anarak area to Kerman – Baluchestan region, and continues to Pakistan, where the Chagai porphyry copper belt is located. There are some well – known porphyry copper deposits in Iranian part of this VPA but certainly there are same grounds to search for new porphyry deposits.

Iran like other developing countries relies on the use of their natural resources to support their economic development. It is clear that, oil is the most important natural resources but, utilization of mineral deposits has a long antiquity in Iran. In particular, the exploration of mineral resources has traditionally been a significant component of the Iran economy.

Increasing of the base metal prices like Copper, Iron, Lead and Zinc especially at the recent years causes to attention for finding new resources more and more because one reason for decline in economic development of many countries is decrease of known mineral deposits.

1-2- Description of the problem

The risks of developing mineral resources need to be known as accurately as possible. This process should start at the pre-discovery stage and continue through feasibility to the development stage. Until recently, this type of analysis was carried out manually and predictor maps were interpreted individually leading to subjective judgments. Nowadays using GIS and mathematical algorithm for digital modeling of predictor maps has supplanted traditional approaches. A program of digital data integration allows using of more probabilistic data analysis techniques at the pre-discovery exploration stage. A GIS was created and spatial analytical techniques employed to assess the potential of the area, and to test current geological models.

In the study area of this research, there are some reports about geology, geophysics, geochemistry, ore deposits… but, many of them were prepared in hard copy format and some others don't have standard database or have disparate datasets, also some datasets are not reliable. Beside that these information haven't considered in regional scale. Therefore development of a geomanagement system using sufficient geosciences data for creation standard datasets that can be useable in GIS environment is vital. In fact a mineral potential map is prepared using combination of mentioned data layers. Although a variety of models and their applications to map mineral potential in local scale are documented in the published literature, but implementation of a comprehensive study involving application of GIS-based numerical methods for mineral prospectivity mapping in this area is essential.

The main problem in spatial-mathematical-model-based approaches to mineral potential mapping is the selection of appropriate functions that can effectively approximate the relation between the target mineral deposits and recognition criteria as well as account for dependencies amongst the recognition criteria. Such a study can help to establish strengths and limitation of different modeling techniques and, therefore can facilitate selecting appropriate model(s) for mapping mineral potential of geologically comparable areas.

1-3- The study area

The study area is located at northwest of Kerman province in central Iran (figure 1-1). This area covers some parts of Sirjan, Rafsanjan, Neyriz, and Anar geological maps with scale 1:250000; and also the main parts of Shahr-e-Babak, Rafsanjan 1, 2, Pariz, and Chahar Gonbad 1:100000 geological maps. The study area is limited between longitudes 55° to 56° 30´ N and latitudes 29º 30´ to 30º 30´ E.

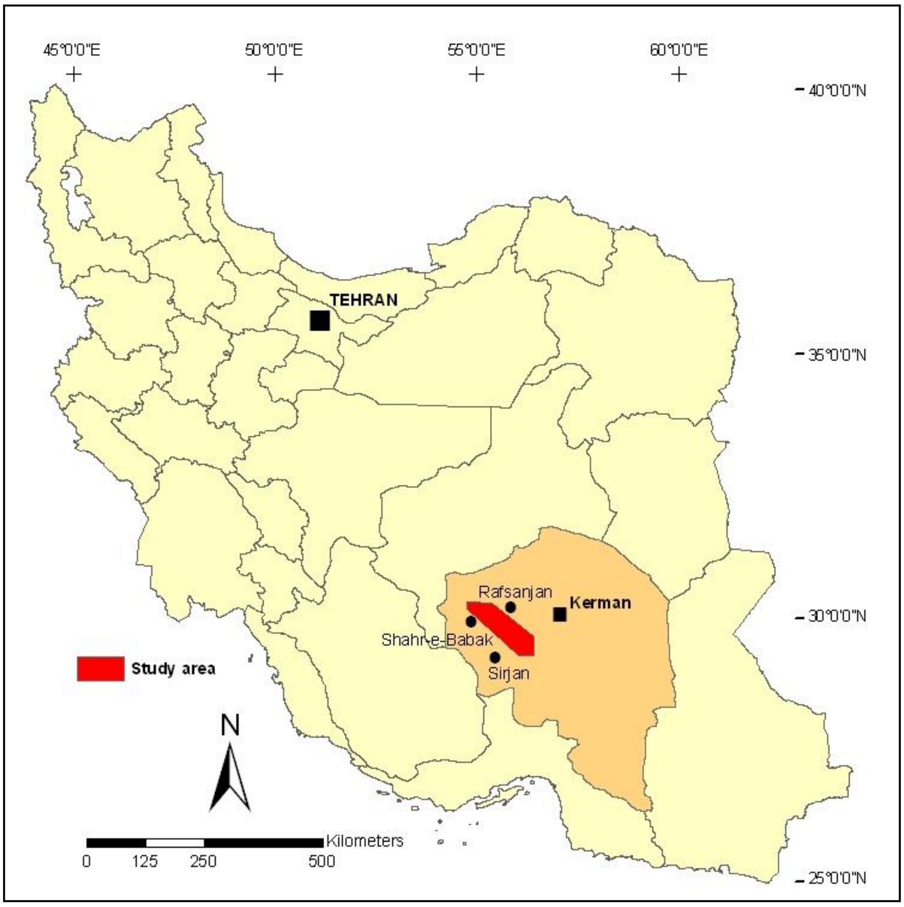


Figure 1-1- The location of study area

2- Geology and Metallogenetic Features of the Study Area

2- 1- Introduction

The tectonic history of Tethyan region has been studied by many authors (e.g. Berberian and King, 1981; Berberian et al., 1982; Jankovic, 1984; Sengor, 1984, 1990, 1991; Alavi, 2004; Shahabpour, 2005; Stampfli and Borel, 2002). From Late Precambrian until Late Paleozoic time, southeastern Turkey, central-east Iran and Arabia were part of Gondwana continent, separated from the Eurasian plate by Paleotethys ocean. The closure of Paleotethys during Upper-Late Triassic time by the northward motion of the Central-East Iran microcontinents resulted in their welding with the Eurasian plate along a suture zone consisting of oceanic rocks. Sometime prior to the Middle Triassic, the Late Paleozoic ophiolites were emplaced in the north, presumably at the time of collision of continental fragments with the Eurasian plate. According to this interpretation, Eurasia appears to have grown from the Middle to

Late Triassic by successive accretion of microcontinents which came from the south, across the Paleotethys ocean. Rifting along the present Zagros fold-and-thrust belt of the continental plate took place in Permian to Triassic time, resulting in the opening of Neotethys ocean.

Interpretations of the formation of Zagros orogeny vary widely, from continental rifting models, to single or double subduction zone models. According to the subduction model, movement of the Iranian plate and expansion of the Neotethys oceanic crust ceased in the early Cretaceous and was then reversed. This resulted first, in the subduction of the Neotethian oceanic lithosphere below the Central-East Iranian Terrane, and subsequently, in collision between the Arabian and Iranian plates. The Sanandaj-Sirjan metamorphic belt is believed to be a continental fragment rifted from the Arabian margin in the Late Permian or Triassic (Hooper et al., 1994; Glennie, 2000). Extrusive volcanism in the Urumieh-Dokhtar magmatic belt began in the Eocene, continuing through the Miocene, with a climax in Middle Eocene (Berberian and King, 1981). It is generally assumed that the Urumieh-Dokhtar magmatic belt was the magmatic arc overlying the slab of Neotethyan oceanic lithosphere which was subducted beneath the Iranian plate. Intrusive activity associated with porphyry copper deposits in the Urumieh-Dokhtar magmatic belt occurred in the Middle Miocene, and appears to represent the final phase of arc magmatism prior to collision (Richards, 2003).

The study area of this research is located in Urumieh-Dokhtar magmatic belt (figure 2-1). This belt is a linear extrusive-intrusive assemblage with a thickness of over 40-50 km trending northwest-southeast. The magmatic assemblage is interpreted to be an Andean Type Magmatic Arc (Alavi, 1980; Berberian, 1981; Shahabpour 1982) or a rift (Amidi, 1984). The petrochemical analysis suggests a calc-alkaline (Forster, 1972; Berberian, 1982) or an alkaline (Amidi, 1984) composition. The belt mainly consists of basaltic lava flows, trachyandesites, trachybasalts, andesites, dacites, trachytes, ignembrites, and pyroclastics. The intrusive rocks are mainly diorite, granodiorite, gabbro, and granite.

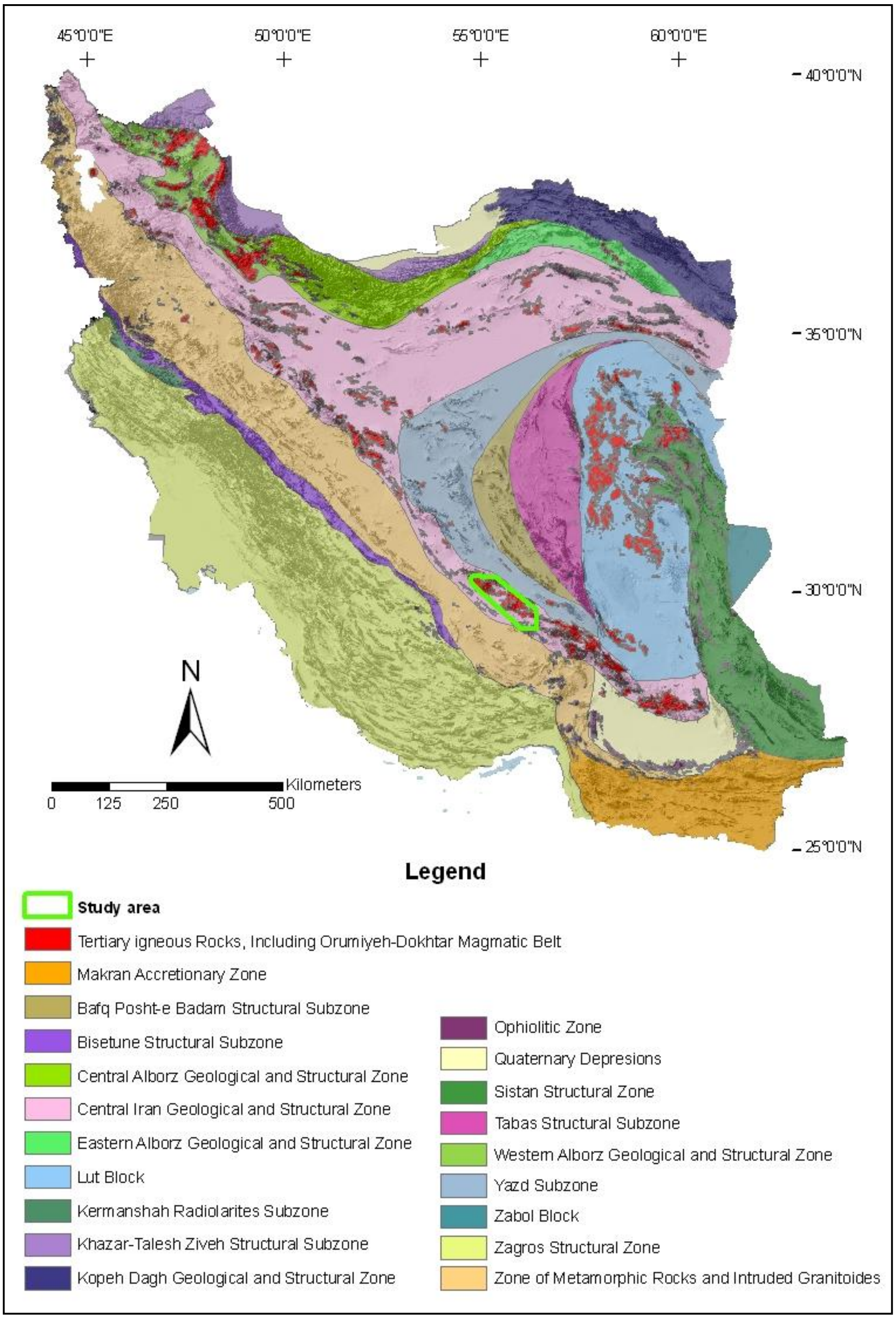


Figure 2-1- Location of study area on structural map of Iran (Sahandi et al., 2008)

2- 2- Geology of study area

The geology of study area may be considered as complex, due to the fact that the different geological formations ranging from the Cretaceous up to the very recent Quaternary sediments. Brief descriptions of several geological units basis on age priority (the youngest up to the oldest) are as follows:

Quaternary

Quaternary sediments consists of dasht blankets deposits between the mountain ridges, clay flats in the main intramountain basins, and sand flats.

Neogene

Neogene deposits appear in several rock types in the study area (figure 2-2). Sedimentary units include sandstone, conglomerate, clay, and siltstone. Volcano-sedimentary units comprise arenites with pebbles and boulders of volcanic rocks, and volcanomictic conglomerate. Volcanic rocks consist of andesite, andesitic-basaltic lava flow, dacite, olivine basalt, and pyroclastics. Intrusive bodies contain diorite, quartz-diorite, monzonite, quartz-monzonite, and granodiorite. Metamorphic units have limited outcrops as hornblende phenoandesite.

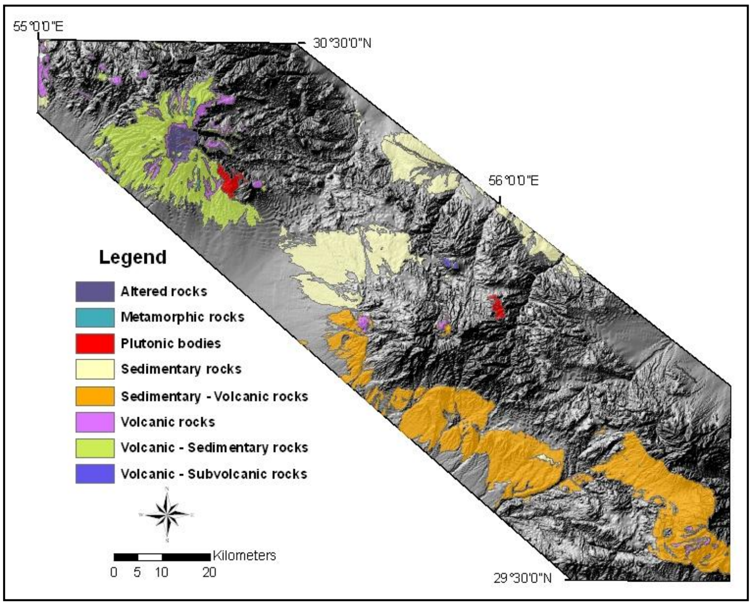


Figure 2-2- Distribution of Neogene units in the study area

Paleogene – Neogene

Paleogene – Neogene rocks have been seen in southern part of study area (figure 2-3). These units include plutonic bodies (garnodiorite and quartz-diorite porphyrite); volcanic rocks (dacite and agglomerate); volcanic-sedimentary rocks (tuffaceous sandstone); and sedimentary units (sandstone, siltstone, limestone, conglomerate, and marl).

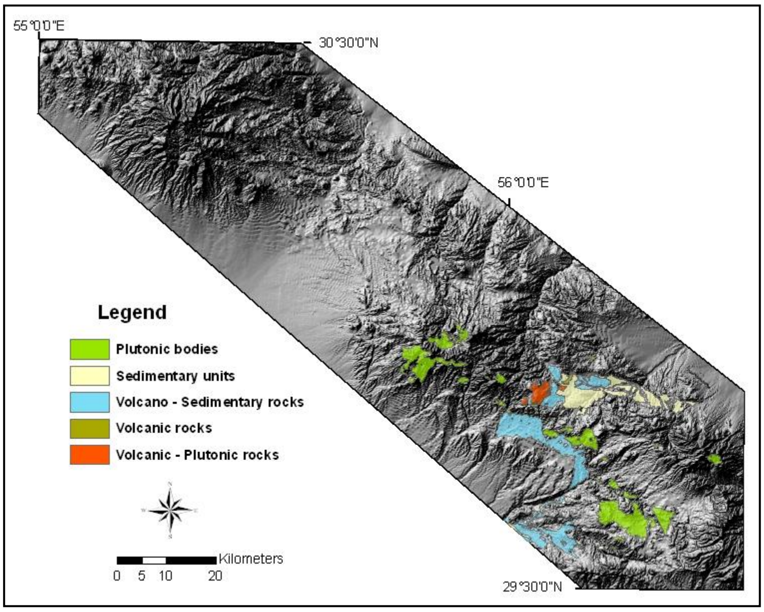


Figure 2-3- Distribution of Paleogene - Neogene units in the study area.

Eocene

Eocene outcrops cover the most parts of study area (figure 2-4). These units comprise plutonic bodies (diorite and diabase), volcanic rocks (trachy-andesite, trachy-basalt, trachyte, andsite, and basalt); volcano-sedimentary rocks (tuffaceous sandstone, volcanoclastics, tuffite, flysch, and tuff); and sedimentary rocks (conglomerate, limestone, marl, and sandstone).

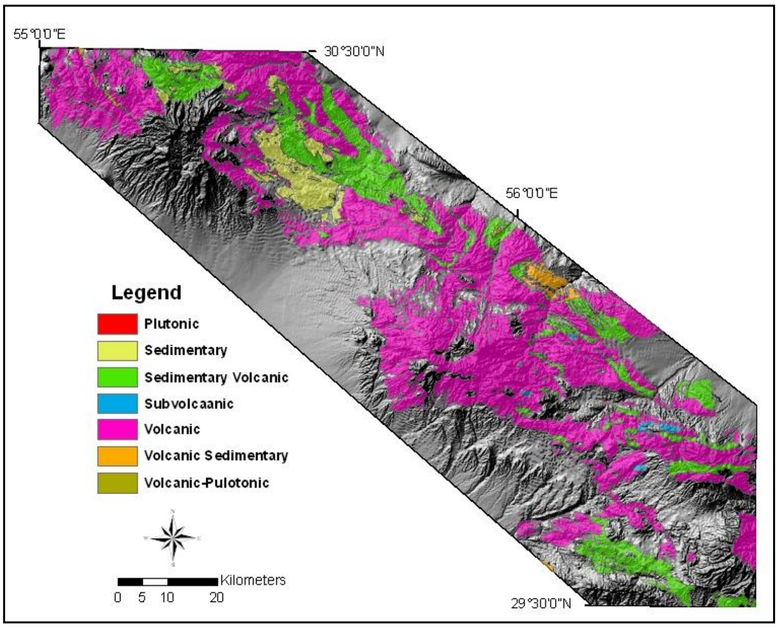


Figure 2-4- Distribution of Eocene units in the study area.

Cretaceous

Cretaceous units can be viewed in southern and northern margin of the study area, dominantly as ophiolite complex. These outcrops contain serpentine and colored melange, diabase and also marl, marly siltstone, and limestone (sedimentary rocks).

2- 3- Metallogenetic features of the study area

The study area is situated in the Alpine-Himalaya orogenic system, where is a part of a huge copper belt, several thousand kilometers long, extending from the East Serbia via Bulgaria, Turkey and Iran, to Afghanistan and Pakistan. Relating to metallogenetic features, of particular importance inside this belt was the magmatic activity, mostly the distribution of ultrabasic and basic rocks, volcanic-sedimentary complex and synorogenic intrusive.

The synorogenic and other intrusive bodies have frequently been the direct carries of porphyry copper mineralization. The copper occurrences and deposits are located in three linear belts:

1. East Serbian – North Anatolian metallogenetic belt.
2. Central Iranian metallogenetic belt.
3. Macedonian – South Anatolian metallogenetic belt.

All three belts represent second-order matallogenetic units in the Alpine-Himalaya planetary metallogenetic belt, and each one has its own characteristic geological, structural, magmatic and metallogenetic features.

The study area of this project is a part of the Central Iranian metallogenetic belt and may be considered as a single metallogenetic zone characterized by similar geological features and by many copper deposits, identical in genesis and age. This metallogenetic zone coincides with the central parts of the Eocene volcanic-sedimentary complex, and extends, in a north-southeast direction, for at least 400 km, with an average width of 40-50 km.

The main characteristics of the mineralization in this belt are as follows:

- They are located mainly in the intermediate Oligocene_Miocene intrusive bodies or Eocene volcanic-sedimentary complex.

- They have been formed as a product of endogene mineralizing processes, in the hydrothermal phase and under different temperatures.
- The fault-type structures, where the mineralization are located or which were of importance in mineral deposition, are mostly of general N-S, NE-SW, and E-W direction.
- The ore-bearing solutions have produced intense hydrothermal alteration of the rocks.
- The mineralization is known to be of Alpine age.

The most significant geotectonic features, related to distribution of mineral deposition in the study area, are the sedimentation, magmatic activity and structural displacement that occurred during the Tertiary. The intrusive complex is of granitic to gabbroid composition, but granodiorite and diorite are the most common rocks. Structural features play an important role both in the spatial distribution of intrusive bodies and in the formation and distribution of mineral deposits and mineralization. The porphyry copper mineralizations are related to regional scale faults (length more than 10 km). The most important trends for mineralization are N-S, NE-SW, E-W, and NW-SE respectively. At places, where the two fault systems intersect, the intrusive bodies are frequently hydrothermally altered. Such a location has the best position for porphyry mineralization (like Sar Cheshmeh mine). Figure 2-5 shows the relation of regional faults and the location of mentioned porphyry copper mineralization in the study area.

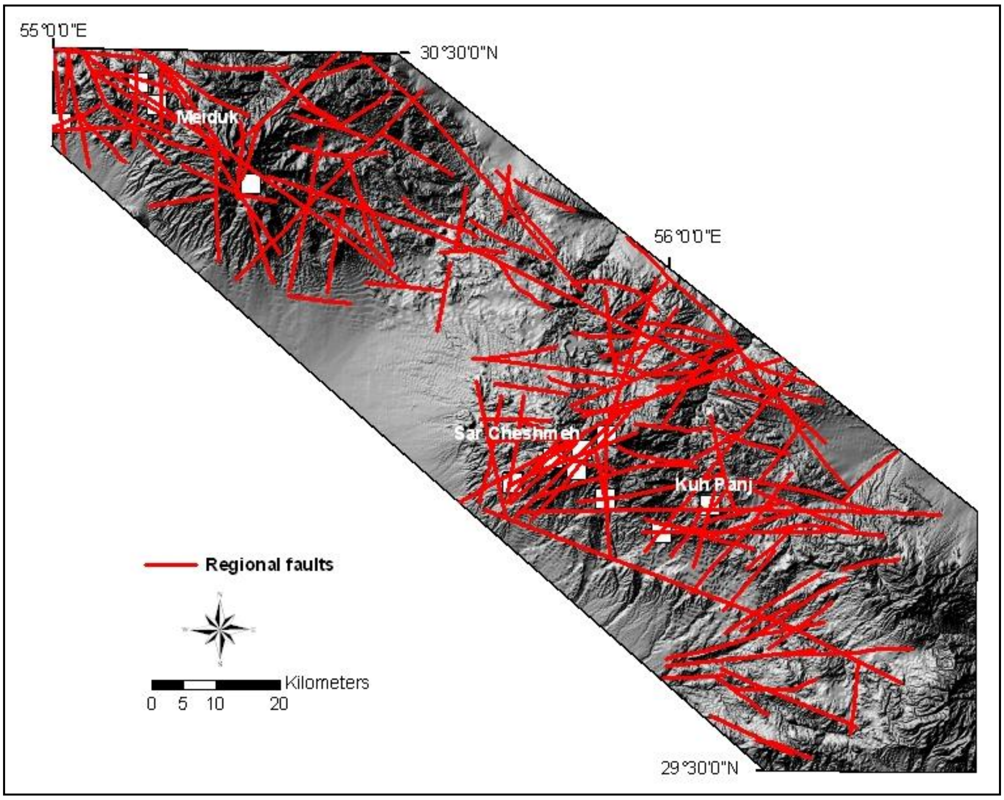


Figure 2-5- Relation of regional faults

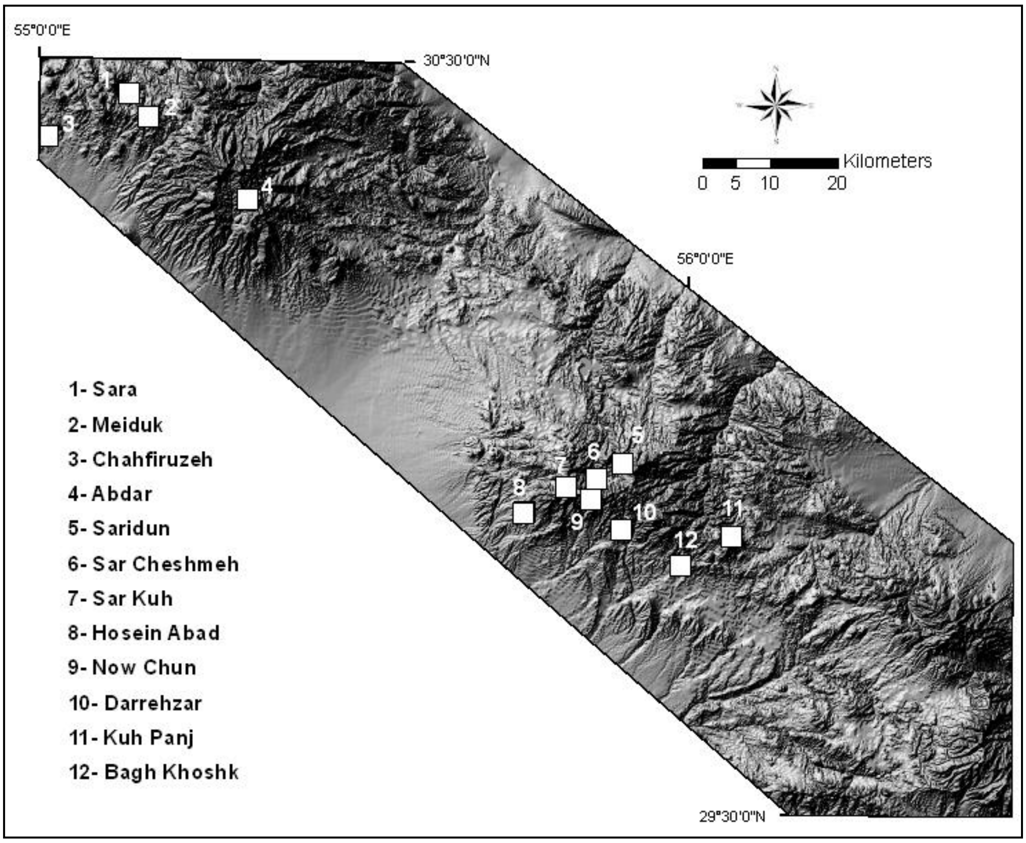


Figure 2-6- Location of porphyry copper mineralization in the study area

3- Geospatial Data Analysis and Modeling

3- 1- Introduction

Geospatial Information Systems present tools to quantitatively analyze and combine georeferenced information from geological, geophysical, geochemical, ore deposit occurrences and remote sensing imagery surveys for decision-making processes. Brilliant coverage of well-documented, accurate and high-quality data enables information production and examination of variable exploration models in an efficient technique with GIS tools.

Mineral-prospectivity mapping is one of the fields in geosciences modeling that is able to take great advantage of using GIS technology as a substitution for traditional and conventional methods. GIS methods have surpassed the human brain in their ability to combine and analyze quantitatively large amounts of geospatial data. There are several mathematical and statistical techniques available for pattern recognition in geospatial data. One of the main advantages using spatial analysis as a prospectivity mapping tool is the capability to store the steps and parameters used in the modeling process, so that the procedure is repeatable and well documented for further analyses.

The methods employed in this research include Artificial Neural Network (ANN) and interval valued fuzzy sets TOPSIS (IVFS TOPSIS) modeling. These methods are based on the applying GIS techniques to preparation of data layers, and also modeling inside (IVFS TOPSIS) or outside (ANN) of GIS environment (MATLAB software).

3- 2- Artificial Neural Network basics

3- 2- 1- Introduction

The word “soft computing” was first raised by professor Zadeh in 1992. He defined “soft computing” as: “a computing technique which is searching some solutions for artificial fulfillment of remarkable capacity of human mind in reasoning and learning in an unspecific and un-exact environment”.

Soft computing is including three techniques of fuzzy logic, artificial neural networks, and genetic algorithms and also their combinations. Each of these has its special features and has the capability of modeling some of human intelligence features.

Neural networks are huge computing models, with parallel structures which are used for saving and processing data. As the name of these networks suggests, their structure is inspired by neural structure and human mind's working method.

3- 2- 2- Artificial neural network

Human mind is the most wonderful existing carbonic computer that its mass is a little more than 1.36 kilograms and nearly contains 10^{11} neural un-dividable cells each of which is celled Neuron. All the human activities and behaviors are reflections of this small unit. Each neuron is in relation with a lot of other neurons, in a way that, this relation between neurons has created a compressed network called neural network.

Artificial neural networks are devices or software that are organized based on mind's neural structure and show some behaviors which similar one, exist in the human mind's function or is interpretable to one of human behaviors.

Considering what was told until now, Kohonen present the following explanation of artificial neural networks:

"Artificial neural networks are completely parallel and joined networks of simple particles (usually comparable) along with sequential structures to be able to reaction with the real world in the same way as biological neural systems".

The history of neural networks returns to year 1943 when the initial model of neuron was introduced.

Now, we state the basic principles which by them, the neural networks can be explained in the mathematics format.

These principles originated from biological world and by utilizing mathematics, we try to explain neuron's biological behavior and their networks.

The basic specification of an artificial neural network is divided to four parts:

A) Network structure: network structure determined that the network is consisting of how many neurons and how these neurons are arranged in the network and how they are connected together. Each neuron or processing element similar to natural neuron has some inputs, synapse power, activation functions, some outputs and bias.

B) Activation functions: activation function specifies the output of a neuron versus of an input.

C) Network's learning algorithm: training algorithm shows the technique of network training for some of specific training pattern.

D) Utilizing methods of artificial neural networks: utilizing method of artificial neural network shows how network response (output) can be specified instead of an input pattern.

In some texts parts B, C and D were explained in one category as operation's features or neurons' dynamic.

Although artificial neural networks are not comparable with natural neural system, have some common characteristics as below description:

- *Learning ability*: Means the ability of arranging network's parameters (synaptic weights) in the time path, that the environment of network experiences new condition, aiming that, if the network was taught for a particular condition and a small change happened in the network's environmental condition (particular situation), the network can also be useful for the new condition with a little training. Also, data in the neural networks is saved in synapses and each neuron in the network is potentially affected by other neurons' total actions. In the result, data are not of separate kind from each other but they are affected by the total network.

- *Data distribution "processing data in the text form":* There is not a one to one relation between inputs and synaptic weights. Each neuron in the network is affected by the total actions of other neurons. In this base if some part of network cells omits or have wrong action, there is the probability to reach to the correct answer.

- *Generalization ability:* The network learns the function, creates the algorithm or finds a suitable analytical relation for the number of dots in the space.

- *Parallel processing:* The cells which are placed in the same level can response to those level inputs synchronically.

- *Resistibility*: The cells correct each other local faults during cooperation process.

3- 2- 3- Neuron physiology

The basic and fundamental unit of neural system, particularly mind, is neuron (neuronal cell). Despite neuron's microscopic size, this small unit is a wonderful factory of processing electrical and biochemical data. In classic view, neuron is a simple processing unit, which receives and combines signals of adjacent neurons, through a set of connected paths. These input paths are called dendrite (figure 3-1).

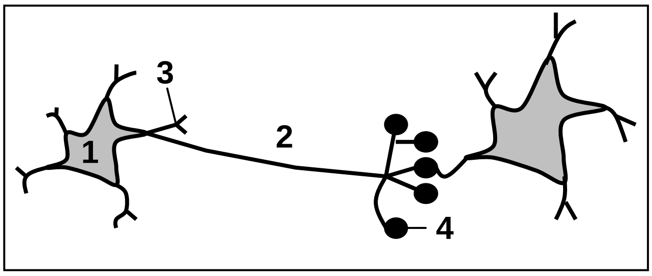


Figure 3-1- Simplified scheme of the neuron:

1- soma, 2- axon, 3- dendrites, 4- synapses (L. Rutkowski 2006)

3- 2- 4- A neuron's mathematical model

Now we concentrate on the basic unit or structural block of artificial neural networks that is neuron (or processing element). Here we mean artificial neuron and should know this neuron is too far from natural biological neuron (it is a very simplified model of natural neuron). A neuron in the general condition has a set of X_j inputs, which number j is changing from 1 to n and shows the source of input signal. Each X_j input, before reaching to the main body of processing unit is weighing by connection resistance or weight factor (for example X_j is multiplied in W_j). Beside that, each neuron has a W_o bias. Also there is an entrance amount for each neuron which neuron shots if neuron's activity reaches or exceeds to that amount (it creates a signal). There is also a non-linear function F , which acts on produced signal R, and an output called O, is produced as the result of function F action on R. O is sent to the other neuron's side through mentioned neuron and is considered as other neurons input. When the neuron is part of a network consist of many neurons, each neuron is called node idiomatically.

In the case that m neurons are in a network, an extra subtitle, i is needed in order to know each single neuron.

Inputs, weights, activity signals, output, threshold and non-linear function sequentially are shown with these signals respectively: و R_i ،W_{ij} ، X_{ij} and F_i.

Mathematical model of a neuron is shown in the figure 3-2.

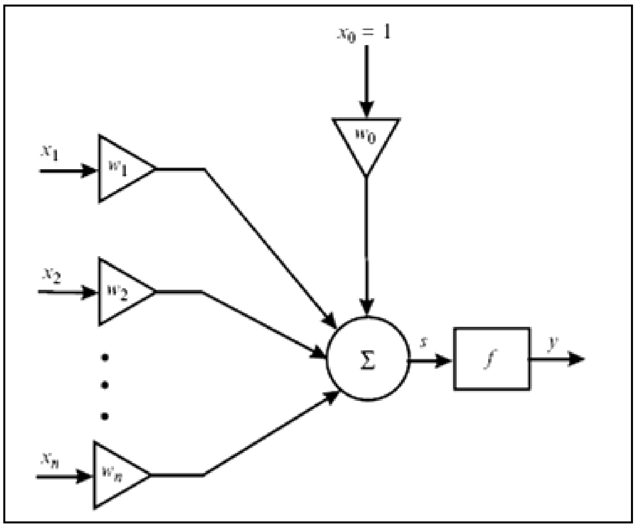


Figure 3-2- Model of a neuron (L. Rutkowski 2006)

Neuron shots in the case this equation is established:

$$\sum_{j=1}^{n} w_{ij} x_{ij} \geq \theta_i$$

Performance function of this basic model is explained as below:

$$y_j = F(\sum x_{ij} w_{ij})$$
$$Y_{in_j} = \sum x_{ij} w_{ij} = x_{0j} w_{0j} + \sum x_{ij} w_{ij}$$
$$= (1) b_j + \sum x_{ij} w_{ij}$$

Where, i shows the number of neuron and j is demonstrating different inputs of other neurons.

3- 2-5- Activation functions

Activation function can be linear or non-linear. The aim of using non-linear function is to be sure that neurons' reaction has some boundaries in front of different stimulations (for example extra big or extra small reactions), in other word is controllable. This matter also conforms to real biological world. Biological neurons limit their reaction intensity in front of stimulations in the same way. But non-linear function which is used in most of neural network samples is not an exact imitation of biological copies, but this function is often used only for the convenience of computing and mathematical modeling.

Because of this, different kinds of non-linear functions are used, which the kind of used non-linear function is dependent to the type of neural network and its training algorithms. Non-linear functions are placed in the specific limit (have boundary). These functions have a high l or a low limit or both of them. Step function and sigmoid function are of most favorite used non-linear functions, because these functions are monotones, have boundary and their derived are simple and at the same time are non-linear functions. Linear functions causes the outputs of network have some amounts out of a definite limit. Here, we mention three cases of activation functions which are examined in this research (figure 3-3).

Sigmoid logarithm function:

$$\log sig(x) = \frac{1}{1+e^{-x}}$$

Sigmoid tangent hyperbolic:

$$\tan sig(x) = \frac{e^{n} - e^{-n}}{e^{n} + e^{-n}}$$

Motive linear function:

$purelin(x) = x$

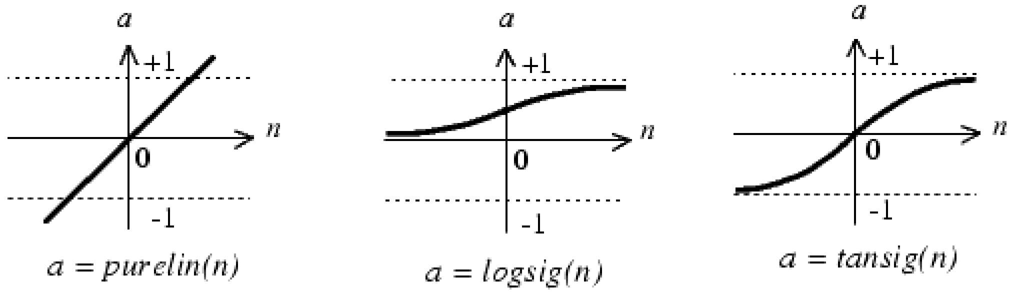


Figure 3-3- Activation functions (Grupal Singh).

3- 2- 6- Neural network's structure

Neurons are naturally connected to each other in a specific way to form a neural network. This placement type on neurons can be in a manner that forms a one layer or multi-layer network. The power of neural computing rises from the quantity of connected neurons in the network structure. Usually, bigger networks present more computing ability. Arranging neurons in different layers or levels is an imitation of a particular part's structure of the human mind. Typically the abilities of multi-layer networks is more than one-layer ones. The most common used neural networks are multi-layer neural networks with training algorithm of back propagation. Here, we explain feed forward multi-layer network.

3- 2- 7- Multi-layer, feed forward networks

Here, we introduce another type of feed forward networks which is known as multi-layer feed forward network. The difference of this network with one-layer network is in this matter that between input layer and output layer there are one or some layers called hidden layers. The responsibility of these layers is connecting the input layer with output layer. Having this hidden layer, the network is able to extract non-linear relations of presented data. Figure 3-4 is demonstration of a three layers feed forward network.

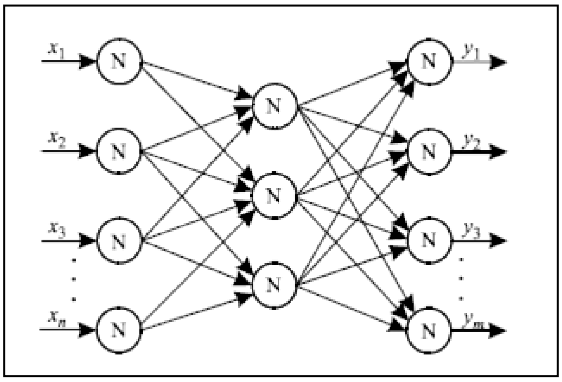


Figure 3-4- A typical scheme of a three-layer feed-forward neural network (L. Rutkowski 2006)

Weights in artificial neural networks take numerical quantity to itself and the network's data is saved in these weights. Each neural network should have a mechanism for training or teaching which can learn the collection of input training vectors, and can regulate the amount of its weights using this mechanism. Input training vectors should contain suitable data about trained subject. A neural network is determined by its structure, training algorithm and processing algorithm.

3- 2- 8- Determination of the best size for network

Freedom degree in an artificial neural network consists of some network's internal connections which are in direct connection to the number of hidden layer's neurons. At the present, there is no way for determination the number of hidden layer's neuron and the only suggested method is the trial and error one. By increasing the number of hidden layer neurons of a small amount to a big one, firstly, the total of errors decrease but after the amount of neurons reaches to the specific number, these errors start to increase.

3- 2- 9- Different training types in artificial neural networks

Artificial network training is a process that, neural network conforms itself for a motive in way that after suitable modification of parameters, network presents desirable response. During training, the network adjusts its parameters (synapse weights), in response to inputs, so that, the real network output converges to ideal output. As a particular training method is suitable for specific persons, for every artificial neural network, a specific method is appropriate as well. Two different types of training are as follows:

3- 2- 9- 1- Training with supervisor

In this training method, network inputs and corresponding outputs are clear beforehand. In support of this network training method, we apply an input to the network. Network in response to that input results the output, this output is compared with the corresponding desirable output. Now, if the real output has discrepancy with the desirable output; the network produces an error signal which is only used for computing the amount of changes that should be done on the synapse weights. This process is done as much as the real output is the desirable output or in close proximity to it. This process for minimizing the errors requires a certain orbit named teacher or supervisor who is considering this point, for adjustment the amount of error, used for the best training. Because of this reason, this training method is called training with supervisor.

Paying attention to this point is important that during training, weights are modified in a way that minimizes errors. In the training time, we may reach to some amounts for weights which apparently make the minimum error in output, but if we continue training process, at first the errors increases and reaches to another minimum again which is less than first minimum. We call that first minimum, local minimum. At this moment if we continue training process and the amount of error does not decrease from the amount of second error, we call this global minimum (figure 3-5). All the training algorithms try to reach to this minimum. To solve this problem, many times should be tried to minimum the error and also each time different primary weights (accidental or non-accidental) to be used.

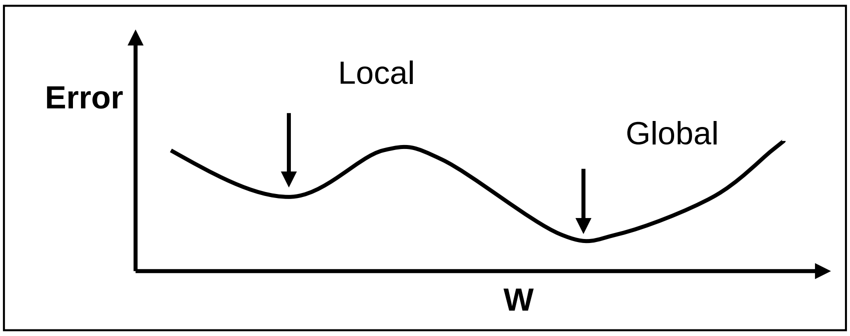


Figure 3-5- Global and local minimum

3- 2- 9- 1- Training without supervisor

In this method in spite of training with supervisor's method, there is no need to teacher; it means there is no aimed output. During training, the network receives its instructional patterns through its inputs and classifies them in different categories desirably. When the network receives an input, shows a response in the output which indicates the category that the input belongs to. If no category is found for it, then a new category is created.

3- 2- 10- Training algorithms

There are three famous laws for neural network's training: Delta rule, back propagation rule or decreasing gradient and Hebbian training rule which in the continuation we only will explain back propagation rule.

3- 2- 11- Training control

To control training of network, different control situations can be used which their most commons will be mentioned here:

- If the number of repetitions exceeds of a specific limit, then training will be stopped (here the intension of repetition is presenting all of the patterns to the network).
- If training speed of α becomes lower than a specific limit, then stop training.
- If amount of mean squared error in the output become lower than a specific limit, then stop training.
- If the changes of all weights become lower than a specific limit, then stop training.

3- 2- 12- Important parameters of artificial neural networks

During artificial neural network's planning, the following points should be considered:

Network structure, number of network's layers, number of neurons in each layer, the type of training algorithms, number of repetition for each pattern during training, number of computing in each repetition, speed of recalling a pattern, network's efficiency, network' flexibility rate (the number of defective neurons and the amount of network efficiency), network capacity or the number of patterns which network can recall, the amount of network compatibility (it means how wide network can adjust itself to new condition after training), bias amounts, threshold amounts, weights limits, selecting activation function, the rate of network security versus noise (it means the number of wrong components in input pattern which network by considering them, can recall ideal output) and remained mode or the final amount of weights that demonstrate the saved program in the network.

3- 2- 13- Separation power of artificial neural networks

Neural networks are taught to separation between input patterns in order to assign for each input, an ideal corresponding response, in the output. In many applications, neural networks are obliged to recognition the pattern and also determination the category which the pattern belongs to. This feature of neural networks is known as separation power, it means the number of patterns which the network can distinguish. Separation power of artificial neural networks depends to its linear or non-linear features which specify a network is linear, multi- linear or non-linear.

3- 2- 14- Specification's extraction

One of the most essential issues in pattern recognition is that, how the patterns' separation features can be extracted from data. Mathematical method for choosing feature is that, we determine common features of patterns and use this features for decreasing pattern vector's dimension.

3- 2- 15- Generalization

One of important characteristics of neural networks is their generalization power. Generalization means the network ability in correct answering to the patterns which has not faced them up to now. Of course the accuracy of network's response is depending on that, the internal relations between input vector components would be same, to the existent relations in training vectors. The ability of finding inner part among training data does not mean well generalization necessarily. A well educated classifying network should response to an experimental data similar to its responses to a training data as well. But if response to the experimental data with lower accuracy, the understanding result is that the network's degree of liberty is not determined in an accurate amount.

Three reasons cause bad results (Romeo 1994): An inaccurate classifying, training algorithm which catches in a local minimum or an inaccurate training collection.

Now we get to know that, how the network can be designed. Designing scheme should give us the possibility of a suitable description of statistical distribution of expansion in data. When the number of input and output neurons became distinct our attention is paid to number of hidden layers and number of neurons inside them. There is no law which can determine the exact number of neurons in the hidden layer. Huang (1991) showed that high limit of required neurons for production of accurate ideal output of training samples is m (the number of training samples). So, the number of neurons in hidden layer should never be more than the number of training samples. Usually the number of training samples should be bigger than the number of internal weights. Practically, $m \approx 10n_{tot}$ is supposed as a good choice. Also the number of neurons should be limited; otherwise there is a risk that taught collection simply is remembered by network (over fitting). Classically, the best arrangement is found by trial and error method which starts by a small numeral of nodes.

Second reason for which the network may not give the ideal reply is response's caught in a local minimum. Error function (misfit) is often too complicated (Hush and others 1992). So, network can simply trapped in a local minimum instead of reaching a global minimum. Finally, the question can be solved using selected training collection. Two problems which often happen are extra training and wrong training. Regarding last case, means selecting wrong amounts of samples (for example assigned patters to wrong category) or training collection does not let a good generalization (for example, general space is not complete). Extra training of studied learning's collection may also involve some problems. Extra training means we imagine the learning collections which are loaded in the memory to be the best for this collection and no other data is acceptable. Increasing in the number of repetition periods also can cause network's divergence. Putting the studied training collection in the memory causes the collection to be adapted with parasite not with the overall answer. In any way for training samples, distribution of weight is gained optimally but generally this problem will have a bad output in total.

Classic solution for this problem is using partition of sample's collections. One part is used for training which is called training group and the next part is used as a validity measuring group for determining total performance limit (figure 3-6). Weights get optimal using utilizing training group. Anyway, validity measuring by second group guarantees a good presentation. When adaptation error of validity measuring group reaches to a minimum; training stops (cross validation holdout method).

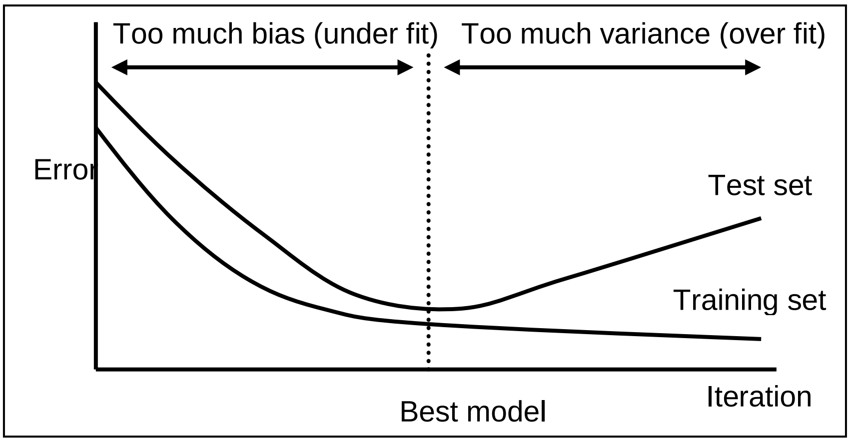


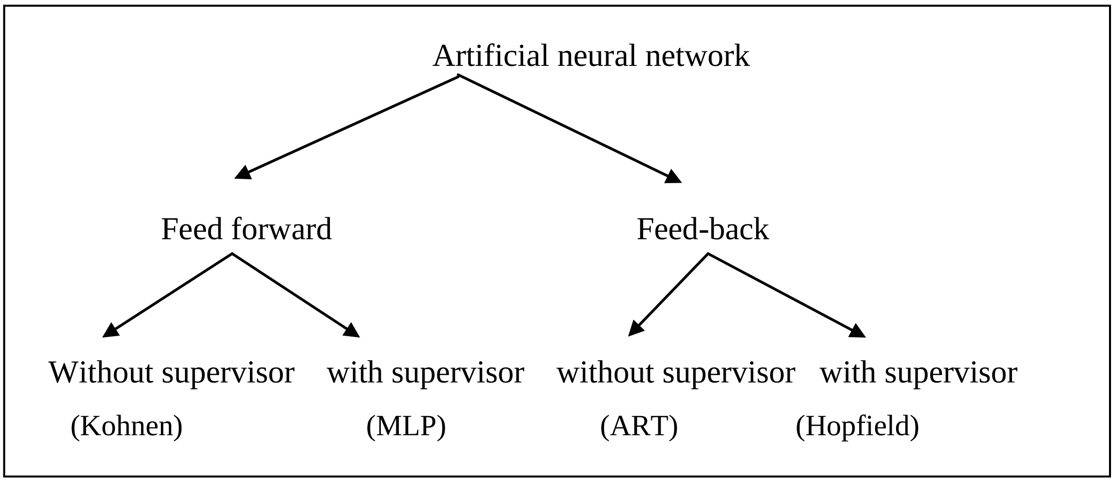
Figure 3-6- Generalization versus training error.

3- 2- 16- Classification of Artificial neural networks

Considering neural network's plentiful variation, these networks can be classified by some of their important characteristics as follow:

- **Type of input:** Neural network's input may be analogue or digital. Networks with digital input usually are used in classifying data applications. But analogue input often related to sound and image processing, recognition and systems' control applications.
- **Training with or without supervisor**: Networks are classified considering the type of network training that was mentioned before.

- **Networks with feed back and feed forward:** Some types of networks have feed back from output to input. This feed forward can be done in different ways. These networks actually have dynamic structures and can be stated by collection of differential equations. This set of neural networks is briefly called reverse networks. In opposite, there is other group of neural networks which devoid of feed forward and do not have any dynamic, that the relation between input and output of these networks can be stated by a static algebraic function. The most applicable artificial networks are with supervisor networks. This kind of networks consist a lot of applications in engineering affairs' scope, because of abundant ability in functions' estimating and non-linear mappings. The most important kinds of with supervisor neural networks, is multi-layer perceptron's networks. Multi-layer perceptron's is a feed-forward network. In the following flowchart a summary of neural network classification can be observed:



For the reason that, multi-layer perceptron network with back propagation training algorithm is the most applicable and the most efficient neural network for solving engineering problem up to now; this method has been used in this thesis. Details of these networks and training method of back propagation will be explained in continuance.

3- 2- 17- Perceptron

Perceptron was the first artificial neural network which was submitted on 1959 by Frank Rosenblatt. He was psychologist and specialist in neuronal biology. Figure 3-7 shows the structure of a perceptron.

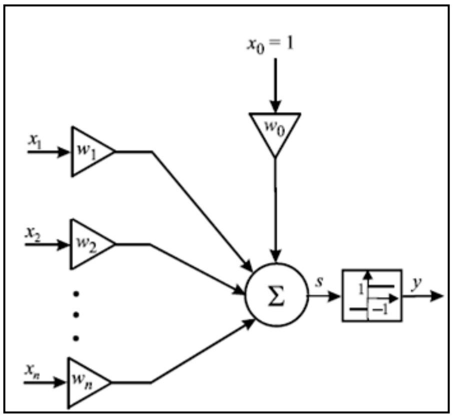


Figure 3-7- Scheme of perceptron (L. Rutkowski 2006)

One layer perceptron with trainable weights is a specific kind of primary perceptron that inside it, activation functions of hidden layer's units are all the same functions. One layer perceptron can classify separable, linear patterns, while multi-layer perceptron can also classify separable, non-linear patterns. Therefore, multi-layer perceptron is used for identifying the pattern and classifying them, due to its power in segregation. For learning multi-layer perceptron, training algorithm called back propagation is used. When Back propagation algorithm is used, all of the activation functions in all of the layers (input layer can have every ideal activation function) should have following characteristics:

a- To be contiguous.
b- Can be derived.
c- To be increasing monotone.

Back propagation algorithm consists of three phases:

1- Entry training inputs and computing network output.
2- Computing the errors and its feed forward from output laver in the direction of input layer.
3- Modifying the weights

The most preferred neural networks' algorithm is multi-layer perceptron network (MLP). Multi-layer perceptron network is a network with layer by layer structure, in such a way that each layer consists of some neurons that their inputs are only connected to the previous layer and their output is connected to the next layer. This network is consisted of one input layer, some hidden layer (usually only one or two hidden layers are used) and an output layer. Hidden layers receives data from previous layer with some weights and after desired operation was done on them; transfers these data with other weights to the next layer up to ends in the output layer.

Sigmoid activation functions are the most common functions used in groups of multi-layer perceptron networks. Outputs of each node in a layer are connected to all of next layer's nodes inputs. Data inside network distributes only in one direction that this direction is from input side to network's output side. For this reason these kind on networks are called also feed forward layer's networks. Existence of more than one layer in perceptron network is the main reason of ability and authority of this kind of neural network. Multi-layer perceptron can do every classification and estimate every non-linear function. Also it is proved that there is no demand to more than three layers for this purpose. It means with three layers perceptron network, every non-linear function can be estimated. It is necessary to remind that to reach to discussed ability; it is exigent that network neurons have non-linear functions (output layer's neurons can be selected linear). In the case that all of network's neurons are linear, simply can show that the network will be the same as a one-layer perceptron network with as much hidden layer as it has.

MLP networks with an output perceptron and a single hidden layer are explained with following formula:

$$f_{MLP}(X) = \sigma \sum_{k=1}^{n_{h1}} w_k \sigma(W^{(k)}.X - \theta^{(k)}) - \theta) \quad (1)$$

where $\sigma(.)$ is activated sigmoid function, X is input, W_k is connection weight of k to output node, n_{h1} is the number of nodes in hidden layer, W^k is the weight of all connections to the k node in hidden layer and θ is the bias. The elements which are printed with capitals show the vector signs. Formula (1) can easily be extended to some output nodes and more hidden layers.

As we will show in the future, total number of input variants n_{tot} which determine the behavior of used neural networks' structure, is important. Fortunately this number is gained easily by equation (1). If n_i indicates the number of input variant, n_{hi} number of neurons in $(i)^{\text{th}}$ of hidden layer and n_0 is the number of output neurons, then n_{tot} for a MLP with a hidden layer by following formula:

$$n_{tot} = (n_i + 1) * n_{h1} + (n_{h1} + 1) * n_0 \quad (2)$$

And for a MLP with two hidden layers is computed with:

$$n_{tot} = (n_i + 1) * n_{h1} + (n_{h1} + 1) * n_{h2} + (n_{h2} + 1) * n_0 \quad (3)$$

As we will see, $\frac{n_{tot}}{m}$ ratio, determines suitable amount of a network optimization where m is the number of training samples.

The aim of training the multi-layer perceptron network is, by finding the best amount for neural network's weights, can establish the functional relation of ideal criteria, based on minimizing network's learning error. The most well-known multi-layer perceptron network's training algorithm is called back propagation method. It is necessary to remind that, mentioned method is training algorithm with supervisor. In fact, back propagation is derived from finding optimum of fastest slope method and is a repetitive algorithm. It means optimal weights of network are gained after some steps of computing reiteration. In the beginning, for training the network, it is necessary to put on a collection of input and output data. This collection forms only a part of favorable function domain, which we are intended to estimate it. Then training algorithm by using this data adapts network's weights so that, for existing inputs in training collection minimizes the difference between network output and desired outputs.

3- 2- 18- Back propagation algorithm

In this section the steps of performing Back propagation algorithm is explained:

- Null step: weight's determination (set the weights equals to accidental small amounts).
- First step: follow steps 2 to 9 until don't gain stop conditions.
- Second step: for each training couple (input and its corresponding output) follow steps 3 to 8.

Feed forward phase (entry inputs to the network and finding their response):

- Third step: each input unit $(X_i, i = 1,2,...,n)$ receives its input signal and distributes it to all of upper units (hidden layer).

- Forth step: every one of hidden layer's units determines its total weighted inputs:

$$(Z_j, j = 1,2,..., p)$$

$$z_{in_j} = v_{0j} + \sum_{i=1}^{n} x_i v_{ij}$$

Then it uses its activation function for determining output:

$$z_j = f(z_{in_j})$$

And it distributes this signal to all of upper layer's units.

- Fifth step: each one of output layer's units $(Y_k, k = 1,..., m)$ computes its total of weighted input:

$$Y_{in_k} = w_{0k} + \sum_{j=1}^{p} z_j w_{jk}$$

Then it uses its activation function for determination of its output:

$$y_k = f(y_{in_k})$$

Back propagation phase:

- Sixth step: each one of output layer's units, receives its aimed amount which is corresponding to input pattern and then, computes error:

$$\delta_k = (t_k - y_k) f'(y_{in_k})$$

Then it computes changes amount of weights and bias which end to it in order to adjustment the weights and bias finally.

$$\Delta w_{jk}(t+1) = \alpha\delta_k Z_j + \mu[w_{jk}(t) - w_{jk}(t-1)] = \alpha\delta_k Z_j + \mu\Delta w_{jk}(t)$$

Where α is training speed, $(\alpha \in [0,1])$ which its duty is to control the error amount in weights adjustment and μ is coefficient of minor changes which controls the amount of weight changes' effect degree in previous step, in weight changes of current step, which is $\mu \in (0.1)$. δ_k consists of error in output of neuron k; and then sending δ_k to all of lower layer's units.

- Seventh step: each one of hidden layer's $(Z_j, j=1,...,p)$, computes total of δ_k inputs of upper layer's:

$$\delta_{in_j} = \sum_{k=1}^{m} \delta_k w_{jk}$$

Then multiples this amount in its derivative activation function in Z_{inj} point to calculate error:

$$\delta_j = \delta_{in_j} f'(Z_{in_j})$$

After that it computes the changes of its weights and bias in order to adjust the final weights and bias.

$$\Delta v_{ij}(t+1) = \alpha\delta_j xi + \mu\Delta v_{ij}(t)$$
$$\Delta v_{0j}(t+1) = \alpha\delta_j + \mu\Delta v_{oj}(t)$$

- Eighth step: every one of output layers adjusts its weights and bias. $(j=0,...,p)$

$$w_{jk}(t+1) = w_{jk}(t) + \Delta w_{jk}(t+1)$$

Each one of hidden layer's adjusts its weights and bias. $(i=0,...,n)$

$$v_{ij}(t+1) = v_{ij}(t) + \Delta v_{ij}(t+1)$$

- Ninth step: if stop conditions are gained (the total of square error becomes less than a definite limit) stop training; otherwise come back to first step.

3- 2- 19- Limitations of Back Propagation algorithm

In order to compute the sensitivities for different layer's neurons in the multi-layer perceptron network, derivation of neuron's conversion function is required. Because of this, such functions should be used that have derivative, in another phrase, derivability of conversion functions is the only limitation of back propagation algorithm. One of common kind of conversion functions which has the above condition, are sigmoid functions. One of these function's characteristic is that, for derivation no necessitate to numerical methods which contain computational errors. In other word, these function's derived can be explained based on the amount of functions themselves.

3- 2- 20- Considerations about Back Propagation algorithm

In this section we announce some of important observations about back propagation learning method which was submitted in previous sections:

- *Production and presentation of training data*

Regarding that, the learning implements basis on submitting collection of input-output training data, attention should be paid on correct collection and representation of these data. In fact during learning process, existence data in learning samples is used for understanding and conclusion of network. So, arrangement of training samples to the network should be in a way that, the network profits of same possibility for learning of all training samples. In more clear word, samples in each learning cycle are used accidentally by same choice possibility in the network.

- *Choosing initial amount*

First step in utilizing back propagation algorithm is determination of initial amounts of multi layer perceptron neural network's parameters. It is clear that a good selection can provide a big assist in faster convergence of algorithm. Therefore, when, there is early information about inputs space and frame, it is better to use network's initial amounts for better option, but in a case that there is no information about input's vector space, it is common to select small amounts accidentally. Wrong selection of multi layer perceptron network's primary parameters, causes the network's loop hold at local minimum, in the beginning. Another defect of primary wrong selection is that, after attainment to a point in time, no remarkable changes will happen in average amount of square error. A simple way for approximation of primary amount of multi layer perceptron network's parameters is the following formula:

Selection of primary amount of neurons parameters $\underline{i^{th}}$ accidentally and non-equal to monotonous distribution in the span of $[-3/N_i^{0.5}, 3/N_i^{0.5}]$ where Ni is the number of $\underline{i^{th}}$ neurons inputs.

- *Low speed of convergence*

As much rate of learning "a" is selected smaller, produced changes in network parameters after each repetition of back propagation algorithm will be smaller. This issue itself causes the movement's path of parameters to be easier; but makes the back propagation algorithm slower. On the contrary, by increasing the length of "a", although learning rate and learning speed of back propagation algorithm increases, but noticeable changes in network parameters from each repetition to next repetition forms, which sometimes causes the instability and swinging of the network. It is called idiomatically, “Network parameters are divergent”.

3- 2- 21- Changing the scale of training data (Normalizing)

One of ways to help training and minimizing network error is normalizing input and output data of the network. The best condition for neural networks is when all of inputs and outputs place in the span of [1, 0] or [-1, 1]. One of reasons for normalizing inputs is that activation functions (such as sigmoid function) cannot differentiate between huge amounts.

For example, a function can easily differentiate between 0.5 and 1 but recognizing amounts of 500 and 1000 is difficult (Bakhshi, 2007).

Data can be normalized in three ways:

A) In this way data are normalized between 1 and -1.

$$pn = \frac{2(p - \min p)}{(\max p - \min p)} - 1 \qquad (4)$$

Where pn is the normalized data, $\max p$ is the maximum amount of p data, and $\min p$ is the minimum amount of p data.

B) In this case base on standard deviation and mean of data, normalizing data takes place that is between 0 and 1.

$$pn = \frac{(p - meanp)}{stdp} \qquad (5)$$

Where pn is normalized data, $meanp$ is the average amount of p data, and $stdp$ is the standard deviation amount of p data.

C) In this way each line’s amounts of data are divided to the maximum amount of that line.

After finishing the step of training network with normalized data, data can be un-normalized. For this, if data is normalized in (A) case, this formula is used:

$$p = 0.5(pn + 1) * (\max p - \min p) + \min p \qquad (6)$$

If the case (B) is used:

$$p = stdp * pn + meanp \qquad (7)$$

And if the way (C) is used, the amount of each line is multiplied in that line's maximum amount of primary data (non-normalized data).

3-3- Applying Artificial Neural Networks method for Mineral Prospectivity Mapping

The progress of more professional and objective Geospatial Information System (GIS) based methods for the combination and analysis of regional exploration geospatial datasets has a potentially significant role in supporting the decision-making process for geologists in field acquisition and selection of exploration targets.

The quantity and quality of regional geospatial data available to the exploration geologist are increasing quickly from a multiplicity of sources such as remote sensing imagery, airborne geophysics and geochemical databases. As the amount of geospatial data increases, so too will the need for an effectual algorithms of integration and analysis the data layers (Cucuzza & Goode 1998).

Artificial neural networks propose a new approach for mineral prospectivity mapping and represent a kind of adaptive computing system that can train from the data. Characteristics of neural networks that make them particularly suitable for pattern recognition and classification consist of their capability to:

a) Extract underlying patterns in a dataset which may be incomprehensible to humans and standard statistical techniques.
b) Function without pre-existing knowledge (e.g. a deposit model).
c) Operate at acceptable accuracy when the data are noisy or contain outliers.
d) Perform well when the input parameters are interdependent and exhibit significant non-linearity.
e) Be flexible and retrained when new data become available.
f) Be used on large varied datasets.

These characteristics imply that neural networks could be compatible to the integration of mineral exploration data layers.

The potential of combining neural network and GIS technologies was recognized by Jeffrey and Rossner (1986) and Ritter *et al*. (1988).

A multilayer perceptron (MLP) neural network is applied for this research because it performs very well for a large variety of problem types and has powerful function approximation capabilities.

The network used in this study is shown in figure 3-8. The network has a 5-5-1 topology (where the numbers refer to input, hidden and output units, respectively). The bias inputs are internal and permanently set to 1.

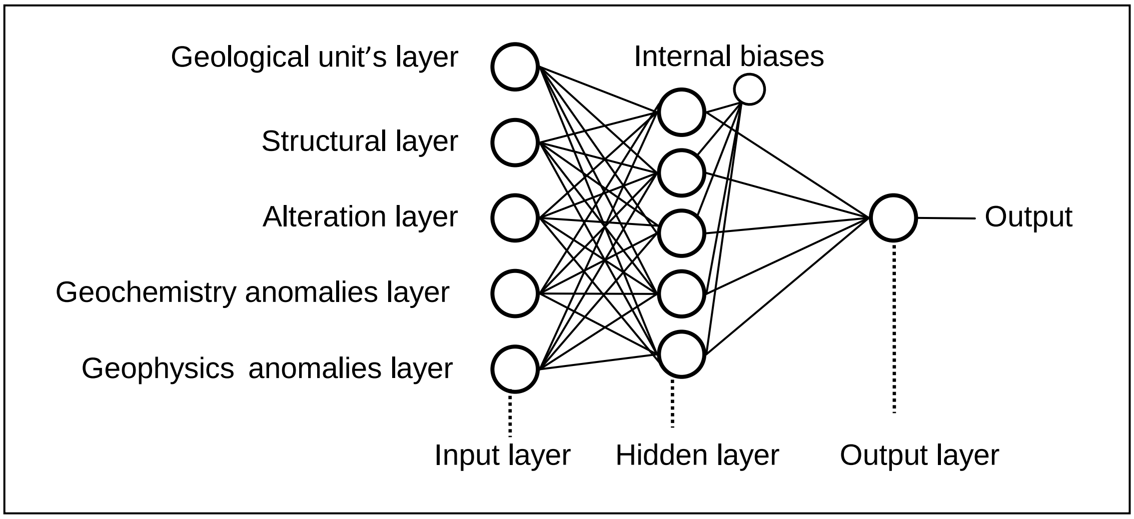


Figure 3-8- Architecture of the multilayer perceptron neural network applied in this study

The neural network consists of input and output layers and one or two hidden layers. Unlike the hidden and output layer units, the input units do not execute any processing and only provide to connect the network to a source of input signals and pass these signals on to other units. Every unit in the hidden and output layers of a multilayer perceptron neural network performs the same series of simple operations involving multiplication and addition, as well as applying a nonlinear transformation to the input signals it receives (figure 3-9). The input on each link is determined by multiplying the input signal by the weight associated with the link. Individual inputs resumed and then transformed by non-linear logistic activation unction (after Hassoun 1995). The output value, referred to as the activation, becomes the signal on all the connections fanning out from the unit. If the unit is in a hidden layer, then this output becomes the input for units at the other end of the links in the next layer.

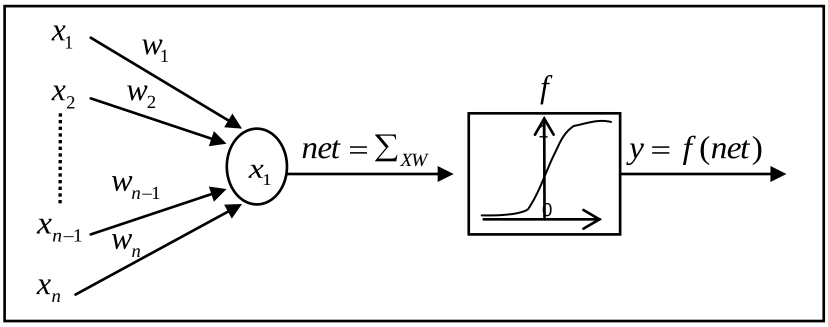


Figure 3-9- Computation's Performance at perceptron unit (Hassoun 1995)

Training in multilayer perceptron neural networks is attained by modifying the connection weights while a set of training examples is repetitively processed. This training method is referred to as supervised because each example specifies all the input parameters together with the desired outputs. The set of inputs to the network constitutes a real-valued vector [x1, . . ., xn]. As the neural network used in this study contains only a single output unit, the training data consist of a list of paired input vectors and scalar output values. The difference between the desired output and the output produced by the network represents the error. Learning continues by applying a learning algorithm (in this research, the back-propagation algorithm) to iteratively adjust the weight values in order to minimize this error. After training is finished, the network should be able to generalize, that is, produce fitting responses for inputs that have not been used during training.

In every training lope, the total sum of squares error follows a decreasing tendency and asymptotically approaches zero as training proceeds. However, excessive learning can lead to what is known as overfitting, related with poor generalization performance. This is due to the existence of noise in the dataset. An excessively trained network can give a very close fit to the training samples, and may be optimal in terms of the total sum of squares error, yet produce very poor generalization. If the number of changeable weights in a multilayer perceptron network is high in relation to the number patterns in the training dataset, and training is continued further than the stage at which the general trends in the data are learnt, the network begins to learn the noise in the dataset. This is harmful to the capability of the network to generalize with new datasets. To avoid overtraining, Masters (1993) recommends a procedure, which involves using (a) a set of training samples to train a neural network, (b) a set of validation samples to validate the trained network and (c) a set of training test samples to check the generalization ability of the trained and validated network. In this study the training samples (278 samples) divided as follow:

a) 70% of samples as training set.
b) 20% of samples as validation set.
c) 10% of samples as test set.

The idea is to improve generalization by stopping learning before the global minimum of the training dataset error, prior to the peculiarities of the dataset are learnt. The training dataset is used to adapt the weights in the network and the test dataset is used to determine when to stop training. The validation dataset plays no role in training and is used to assess the generalization performance of the trained network when new data are processed. Although the training error decreases during training, the test error decreases to a minimum and then begins to increase as training continues. Training is stopped at the point at which the test total sum of squares error reaches a minimum value and the weights corresponding to the optimum number of cycles are used for all subsequent processing.

3-3-1- Data preprocessing

Data processing in the neural-network method includes of three main steps: a) pre-processing of data to production an appropriate format as inputs to the neural network; b) training and processing the whole dataset after training; and c) post-processing for converting the output values to a georeferenced map.

Neural networks, can apply only numeric input data, and performs better if numeric input data are in a fairly-narrow range. This causes a problem when input data are non-numeric or missing or if they are in a wide range. Real-world geospatial data are often categorical (non-numeric) and commonly missing in some parts.

Categorical predictor maps can be converted into numeric predictor maps using one-of-n encoding schemes (Masters, 1993), which involve transforming a multi-class categorical predictor map containing n classes to n classes numerical predictor maps. Therefore all thematic layers in the GIS database were converted to raster format prior to further processing and then reclassified to the range [1, 9]. Each cell in raster format (grid data) present 60 m * 60 m on the ground. There are a total 1875 cells in x-direction (columns) and 2399 cells in y-direction (rows) giving a sum of 4498125 cells.

3-3-2- Create thematic data layers

The thematic layers (predictor maps) were performed, on the basis of a primary conceptual-genetic model. These maps consisting of five layers as follows:

- ***Geological unit's layer***

The original geology map contained 80 different lithology and 1344 polygons as presented in table 3-1.

Table 3-1- The number of different lithology and polygons in geology map.

Numb	Lithology	Total polygons of each	Numb	Lithology	Total polygons of each
1	Q1	214	41	Ettb	6
2	Qf	124	42	Ngpy	6
3	Q2	108	43	2OL,Mag	5
4	Ev	64	44	Ng2c	5
5	Ec	60	45	1OL,M	4
6	Ng	57	46	3OL,M	4
7	Ea	44	47	K2-2,3d	4
8	Ol,Mgd	41	48	K2c	4
9	EV	34	49	Qls	4
10	Edt	34	50	Qtc	4
11	Qal	33	51	B	4
12	Etd	30	52	2Ef	3
13	E_t	29	53	El	3
14	Ng2	26	54	Eta	3
15	Nga	23	55	Eab_t	2
16	Ng1	22	56	Eabt	2
17	Evs	20	57	Ed	2
18	Ngvc	20	58	Es	2
19	3Ef	18	59	Et	2
20	Eab	18	60	Etr	2
21	Nga,b	18	61	K2-2,3CM	2
22	dc-a	18	62	Ngaa	2
23	K2-2,3s	16	63	Ngrs	2
24	K2-2,3l	15	64	Ol	2
25	Ol,M	15	65	D	2
26	Ers	14	66	E,Ol	1
27	Ngp	14	67	Eas	1
28	Ol,Ml	14	68	Esl	1
29	2OL,M	13	69	Etab	1
30	Els	13	70	Etb	1
31	Ha	13	71	Ete	1
32	Qd	13	72	Ett	1
33	1Ef	10	73	K2m	1
34	Eat	10	74	Ng1t	1
35	Qsf	9	75	Ngar	1
36	Ets	8	76	Ngs	1
37	Dp	8	77	Dc	1
38	Elm	7	78	M	1
39	Ol,Mdc	7	79	Qd	1
40	Qt	7	80	Qm	1
SUM			1344		

Basis on the rock types and periods these geological units were classified to 32 groups and then were reclassified to 6 categorical and 6 numerical classes (table 3-2).

Table 3-2- Classification and reclassification of geological units basis on the rock types and periods.

ROCK_TYPE	PERIOD	Categorical value	Numerical value
Sub-volcanic	Eocene	A	9
Plutonic	Neogene	A	9
Plutonic	Oligocene-Miocene	A	9
Sub-volcanic	Paleogene	A	9
Plutonic	Paleogene	A	9
Volcanic Plutonic	Paleogene-Neogene	A	9
Plutonic	Paleogene-Neogene	A	9
Volcanic Plutonic	Paleogene-Neogene	A	9
Plutonic	Cretaceous	B	5
Volcanic	Eocene	B	5
Volcanic Plutonic	Eocene	B	5
Volcanic	Neogene	B	5
Volcanic	Neogene	B	5
Altered rocks	Neogene	B	5
Volcanic-Sub volcanic	Neogene	B	5
Volcanic	Paleogene	B	5
Volcanic	Paleogene-Neogene	B	5
Volcanic Sedimentary	Eocene	C	3
Volcanic Sedimentary	Neogene	C	3
Sedimentary Volcanic	Eocene	D	2
Sedimentary Volcanic	Neogene	D	2
Sedimentary Volcanic	Paleogene	D	2
Sedimentary Volcanic	Paleogene-Neogene	D	2
Sedimentary	Cretaceous	E	1
Sedimentary	Eocene	E	1
Sedimentary	Neogene	E	1
Gaz Gerd formation	Paleocene	E	1
Sedimentary	Paleogene	E	1
Sedimentary	Paleogene-Neogene	E	1
Sedimentary	Quaternary	E	1
Metamorphic	Neogene	F	1
Ophiolite	Cretaceous	G	1

Figure 3-10 shows the categorical classes of geology map in the study area.

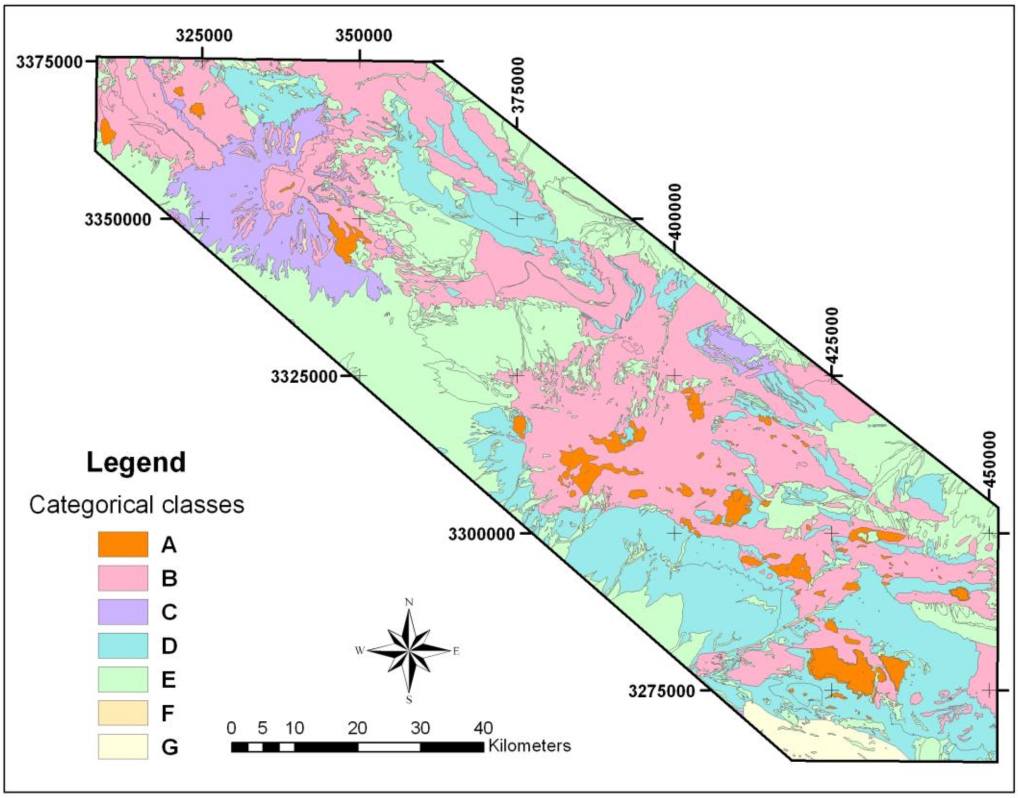


Figure 3-10- The classified categorical geology map of study area

- ***Structural layer***

The structure features (faults) were extracted from 3 different sources: a) geological maps contain 550 lineaments; b) remote sensing imagery contains 154 lineaments; and c) geophysics data contain 423 lineaments. Basis on the porphyry copper mineralization model, the regional faults (with length more than 5 Kilometer) were selected (consisting of 189 lineaments) and then, after calculating the azimuth, were divided to different trending. Final selected faults were buffered as multiple rings in 500 meter and 1000 meter. According to this classification, 7 categorical values, subsequently 7 numerical values were assigned to structure layer (table 3-3).

Table 3-3- Categorical and numerical values, which were assigned to structure layer

Trending	Buffering distance (m)	Categorical value	Numerical value
N – NE	500	A1	8
N – NE	1000	A2	6
E – W	500	B1	5
E – W	1000	B2	4
NW – SE	500	C1	3
NW – SE	1000	C2	2
Background	0	D	1

Figure 3-11 shows the categorical classes of structure map in the study area.

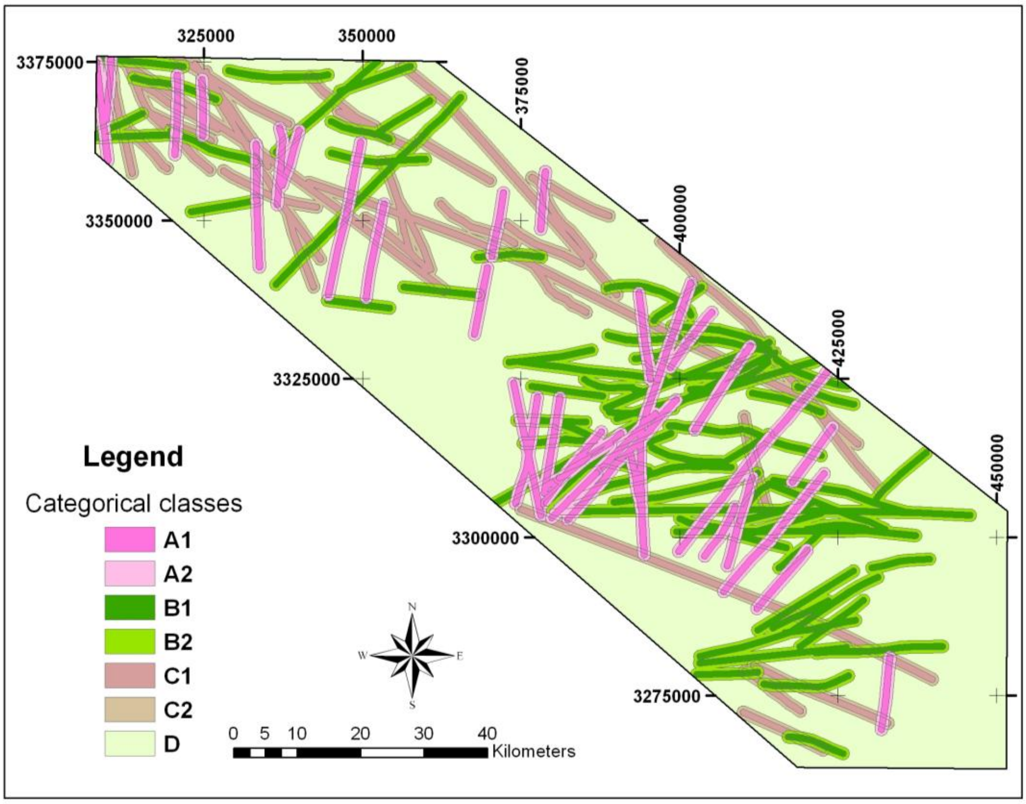


Figure 3-11- The classified categorical structure map of study area

- ***Alteration layer***

Basis on the satellite imagery interpretation (as it was described before) different alteration areas were determined. These areas include four type of alteration that three ones are more important for porphyry copper mineralization (phyllic, argillic and propylitic). The sum of altered areas is 266 square kilometer that consists about 3.5 percent of the study area. These area bases on the conceptual model of porphyry copper mineralization classified to 5 categorical classes, subsequently reclassified to 5 numerical classes (table 3-4).

Table 3- 4- Categorical and numerical values, which were assigned to alteration layer

Characteristics	Categorical value	Numerical value
Zonation of Porphyry copper alteration	A	9
Phyllic alteration	B	6
Argillic alteration	C	4
Propylitic alteration	D	2
Background	E	1

Figure 3-12 shows the categorical classes of alteration map in the study area.

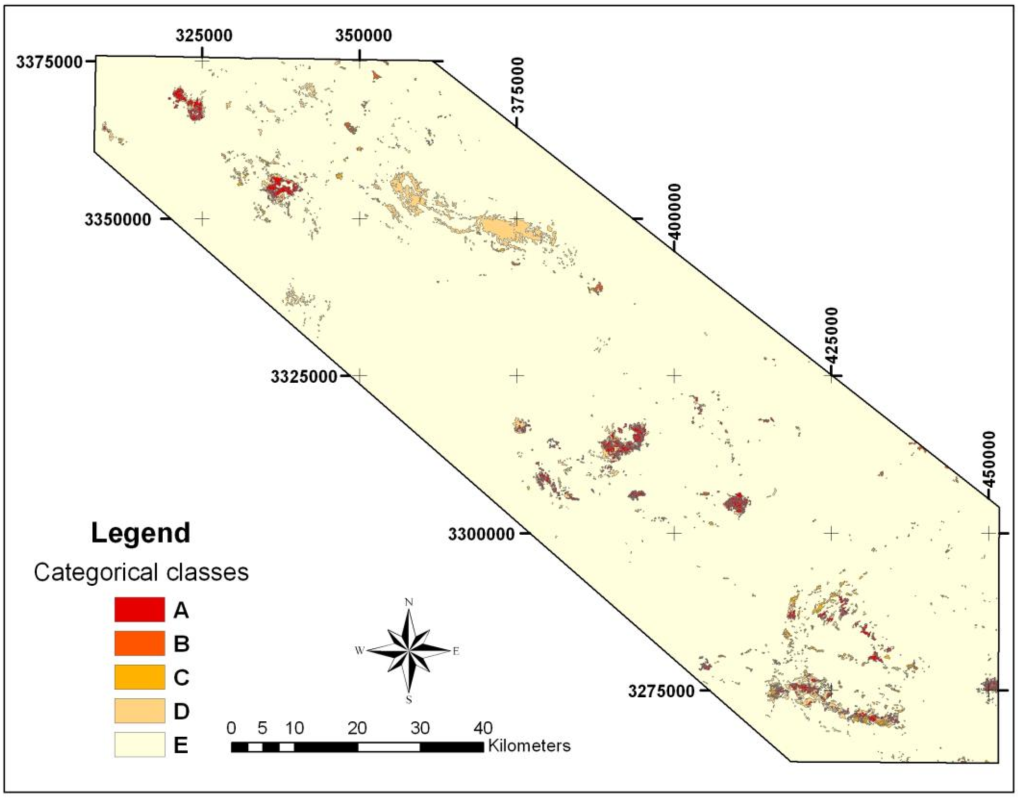


Figure 3-12- The classified categorical alteration map of study area

- ***Geochemistry anomalies layer***

As it was notified before, several geochemical anomalies were found out during geochemical stream samples and heavy mineral processing and analysis. The sum area of these anomalies (79 polygons) is 459 square kilometer that involves 6 percent of the study area. These anomalies basis on the zonation of paragenesis element (Cu, Mo, Pb and Zn) and also content of copper classified to 5 categorical classes, subsequently reclassified to 5 numerical values (table 3-5).

Table 3-5- Categorical and numerical values, which were assigned to geochemistry layer

Characteristics	Categorical value	Numerical value
Show zonation	A	8
High Cu	B	6
Medium Cu	C	4
Low Cu	D	2
Background	E	1

Figure 3-13 shows the categorical classes of geochemical map in the study area.

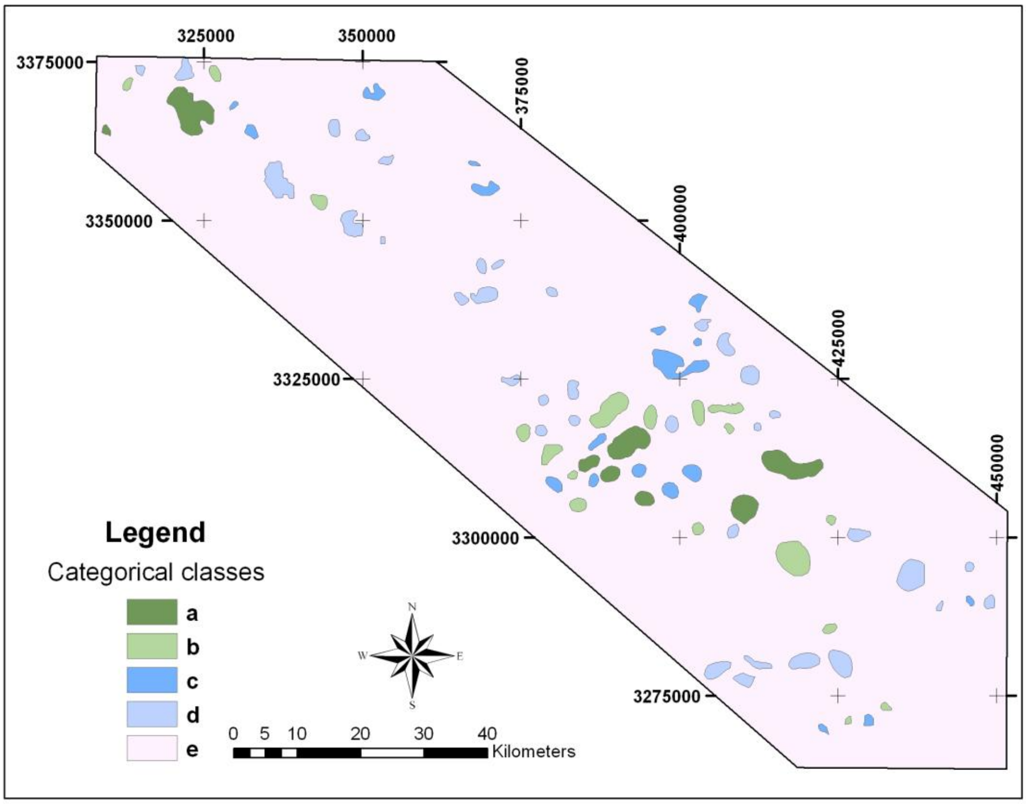


Figure 3-13- The classified categorical geochemical anomalies map

- ***Geophysics anomalies layer***

The outcomes of geophysical data interpretation concluded three types of results: a) intrusive bodies; b) alteration areas; and c) lineaments (were utilized in structural data layer). The intrusive bodies contain 13 districts that, with 181 square kilometer, cover 2.4 percent of study area. The alteration zones include 136 regions that envelop 5 percent of study area. These anomalies were classified to 4 categorical classes basis on the existence of zonation (altered area bordering with intrusive bodies), and also the content of data intensity and the correlation of anomalies with geological features. Afterward, they reclassified to 4 numerical values (table 3-6).

Table 3-6- Categorical and numerical values, which were assigned to geophysics layer

Characteristics	Categorical value	Numerical value
Alteration adjacent intrusion (zonation)	A	9
Alteration with high intensity and correlation	B	5
Alteration with medium intensity and correlation	C	2
Background	D	1

Figure 3-14 shows the categorical classes of geophysical anomalies map of study area.

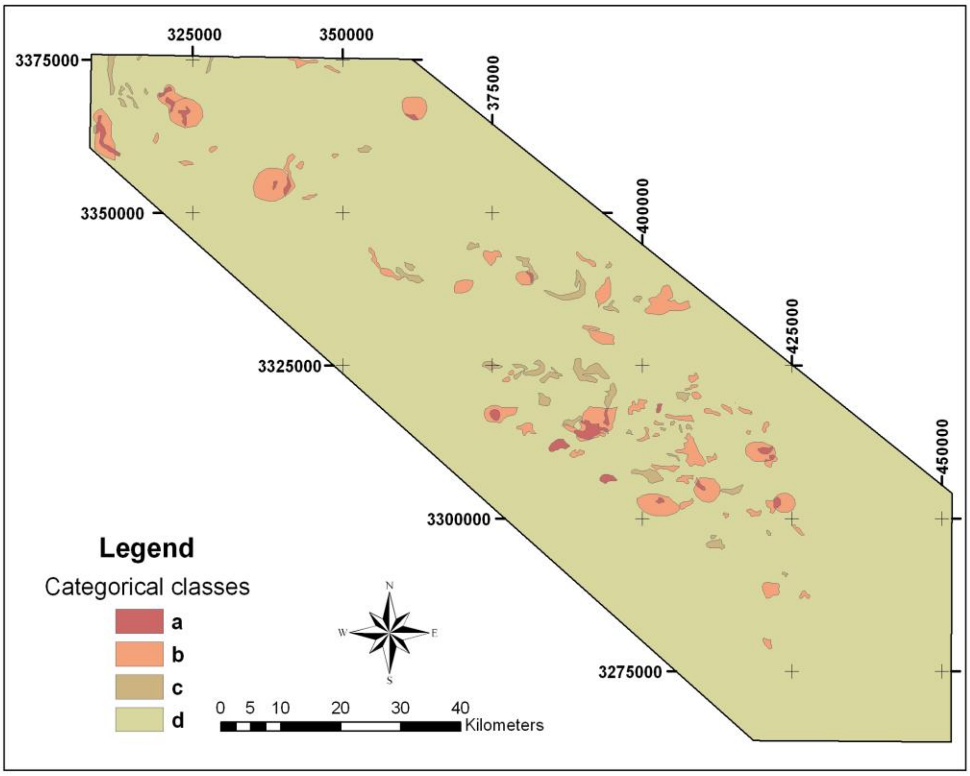
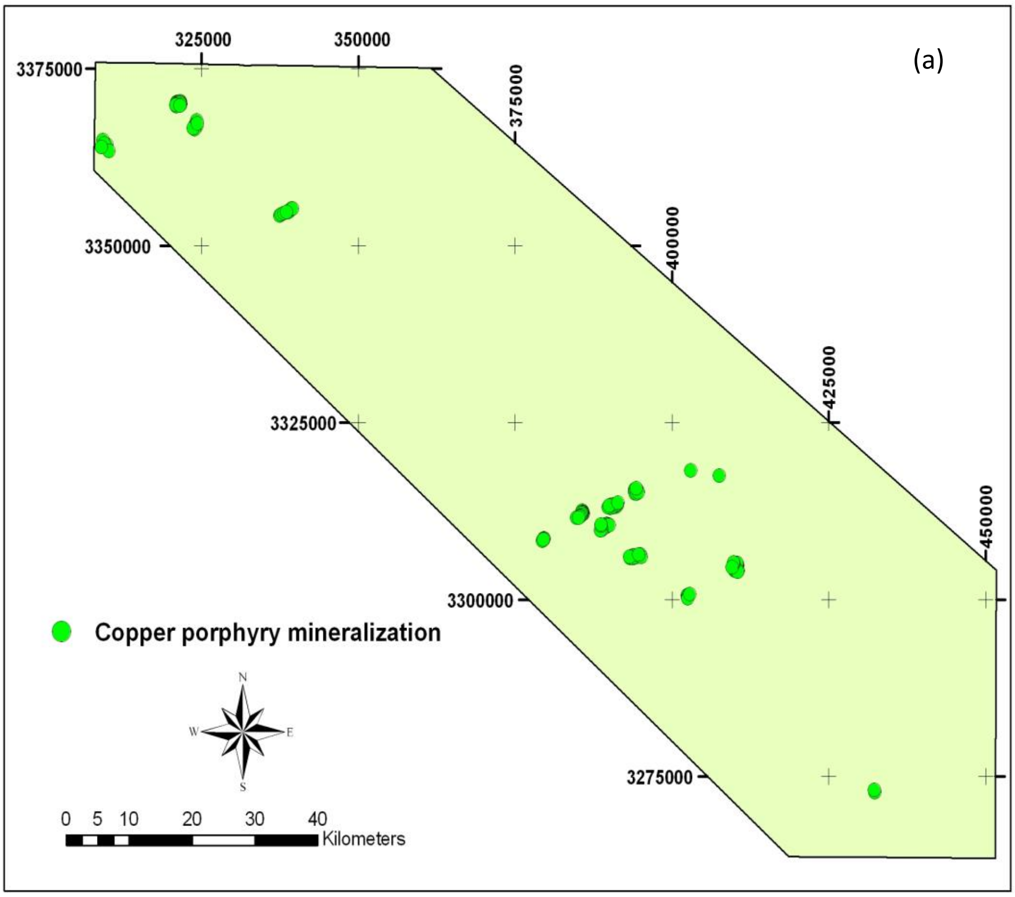


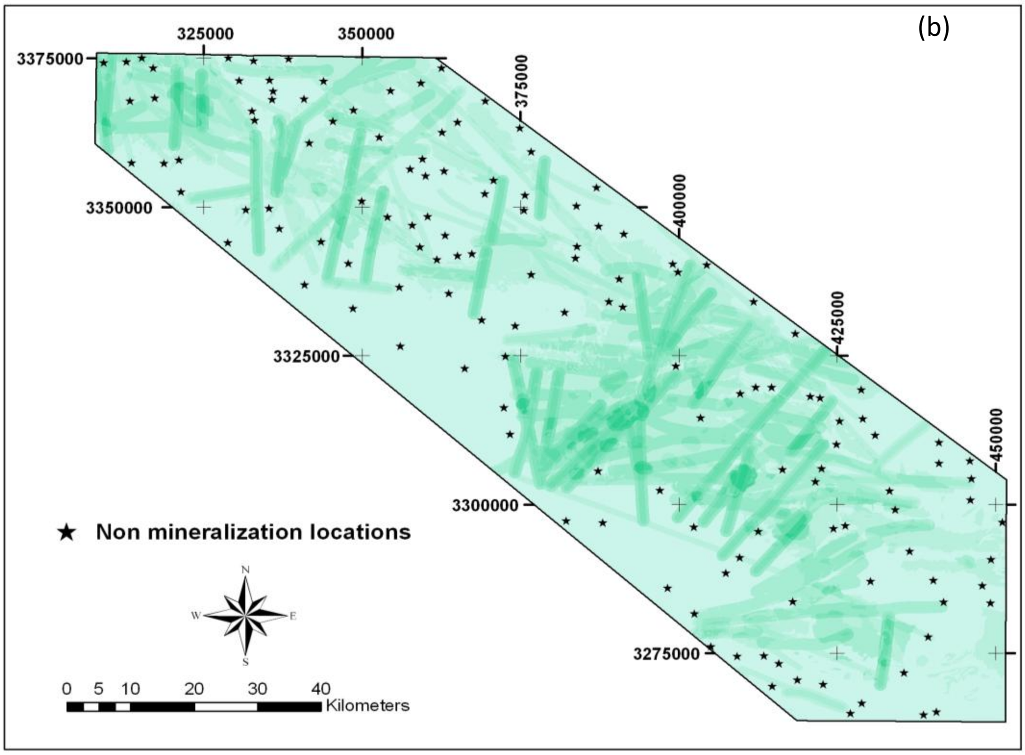
Figure 3-14- The classified categorical geophysical anomalies map

3-3-3- Pattern recognition for training

Network for training needs a pattern, that demonstrate which cells belong to the most positive (assign value 1) and which ones belong to the most negative (assign value -1) locations (copper mineralization point of view) and what in the value of these cells in raster datasets. For this reason, at the first step, 278 locations were selected; contain 139 deposits and 139 non deposits points. The deposit locations were mapped during field surveying using GPS. These points consist of the active porphyry copper mines or confirmed porphyry copper mineralization (figure 3-15a). The non deposit locations were chose basis on the geological evidences that implied to the most inappropriate places for mineralization. For correct decision, these locations were made sure with a simple overlaying of data layers (figure 3-15b).



Figures 3-15a- Location of deposits in the study area



Figures 3-15b- Location of non-deposits in the study area

The next step for making the training pattern is creation a table that shows the values of cells from a set of raster datasets (geospatial data layers with numerical value) for defined locations (deposit and non deposit points). Figure 3-16 shows a schematic analysis of this process.

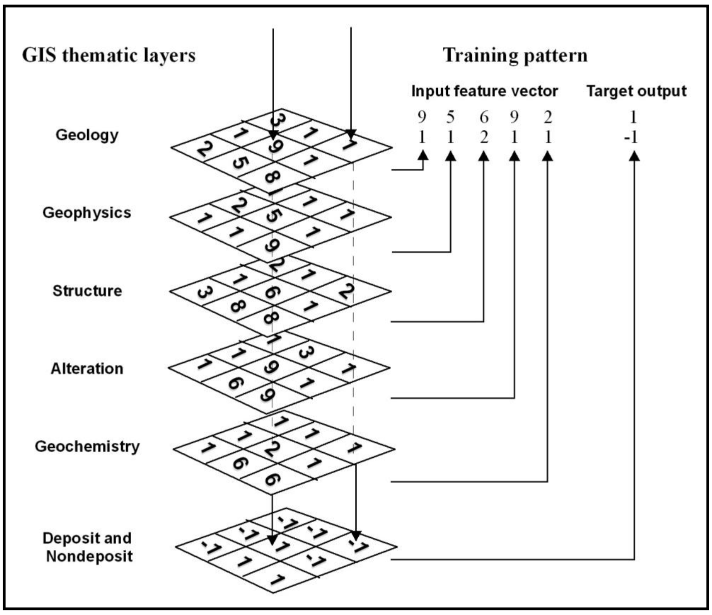


Figure 3-16- Extracting training pattern, including input feature vectors and target output

Following table (3-7) shows some records that are concluded as a result of applying the mentioned procedure. The complete table include of 278 locations.

Table 3-7- Extracted table for pattern recognition

Number	Fault	Geology	Geophysics	Alteration	Geochemistry	CODE
1	2	1	1	1	1	-1
2	1	1	1	1	1	-1
3	1	1	1	1	1	-1
4	1	1	1	1	1	-1
5	1	1	1	1	1	-1
6	1	1	1	1	1	-1
39	2	1	1	1	1	-1
40	2	1	1	1	1	-1
41	1	1	1	1	1	-1
42	2	1	1	1	1	-1
43	1	5	1	1	1	-1
44	1	1	1	1	1	-1
45	1	1	1	1	1	-1
46	1	1	1	1	1	-1
47	1	1	1	1	1	-1
157	1	9	9	9	2	1
158	1	9	9	9	2	1
161	8	9	9	9	8	1
162	8	9	9	9	8	1
163	8	9	9	9	8	1
164	8	5	1	9	4	1
165	8	5	1	9	4	1
170	8	9	5	1	8	1
231	6	9	9	1	8	1
232	8	9	9	9	8	1
233	6	9	9	1	8	1
234	8	9	5	1	8	1
235	8	9	5	1	8	1
236	8	9	1	9	8	1
237	6	9	5	9	8	1
238	8	9	5	1	8	1
276	8	5	1	1	6	1
277	8	9	9	6	6	1
278	4	9	1	9	6	1

Further processing (training the network and modeling with neural network algorithm) is carried out outside the GIS environment using MATLAB software.

3-3-4- Training network using training pattern

For training the network, a program is published in MATLAB, basis on the following steps:

1- Definition of primary error content.

2- Reading extracted table (xls file) as input data and localized it in a matrix (p).
3- Arrange the input data accidentally for abstention of incorrect training (over training).
4- Dividing input datasets to three series as following manner:
    a. 70 percent of input datasets as training data.
    b. 20 percent of input datasets as validation data.
    c. 10 percent of input datasets as test data.
5- Training the network using a “for” loop for determination of hidden layer’s neurons. This loop trains the network with “k” number of neuron (k=1:12), and then compare the results basis on the error content.
6- Introduction the activation functions for a network with “k” neurons in hidden layer and one output. In this research two activation functions (“logsig” and “tansig”), with Levenberg-Marquaredt’s training algorithm were employed. Regarding to results, implied that, using “logsis” activation function culminated in better outcomes (with less error content). Therefore, this activation function is utilized for network’s training.
7- Allocation the primary values to weights and bias of network.
8- Beginning the training process basis on the defined parameters.
9- Simulation of 278 data table using network.
10- Calculation of network’s error with comparison of network’s output and real output.
11- Calculation of proportional error for each data (278 dataset).
12- Creation a matrix (er) that the rows specify the number of neurons (12 rows), and three columns indicate respectively: mean proportional error of training data for defined neuron number, mean proportional error of validation data for defined neuron number, mean proportional error of test data for defined neuron number.
13- Selection of the best neuron number base on the least mean proportional error content and saving this network.
14- Storing weights and bias in w1, w2, b1 and b2 variants respectively.

As mentioned in above paragraphs, determination of mean proportional error of test data series has a main role for selection the best network topology. Hence, in the following sentences, more detailed issues were described.

7- 2- 20-2- 1- Calculation of mean proportional error for test data series.

Determination of mean proportional error for test data series is calculated as follow:

$$re(j) = \frac{100}{n_{tst}} \sum_{i} \frac{|t(i,j) - a(i,j)|}{t(i,j)} \qquad (8)$$

$$i = 1, \ldots\ldots, n_{tst}$$

$$j = 1, \ldots\ldots, n_p$$

Where, $re(j)$ represent mean proportional error for j th parameter in test data series; n_{tst} is the number of test data series; t is the real value of target in i th of test data for j th parameter; a is calculated error value (after simulation) by network, for j th parameter, in i th of test data series.

RMS error is determined for each parameter in test datasets and it is demonstrated for the functionality of network. This error is calculated using following equation:

$$RMS = \left[\frac{1}{n_{tst}.n_p} \sum_{i,j} \left(\frac{t(i,j) - a(i,j)}{\max(j) - \min(j)} \right)^2 \right]^{\frac{1}{2}} \qquad (9)$$

$$i = 1, \ldots\ldots, n_{tst}$$

$$j = 1, \ldots\ldots, n_p$$

Where, n_{tst} is the number of test data series; n_p is the output's parameters (the number of output's neuron); t the real value of target in i th of test data for j th parameter; a is calculated error value (after simulation) by network, for j th parameter, in i th of test data series, and $\min(j)$ represents the minimum value in i th parameter of test data series.

After error calculation, and analysis the results, it is cleared that, the network with 5 neurons in hidden layer has the least error (0.00014); therefore this topology is selected for network (figure 3-17).

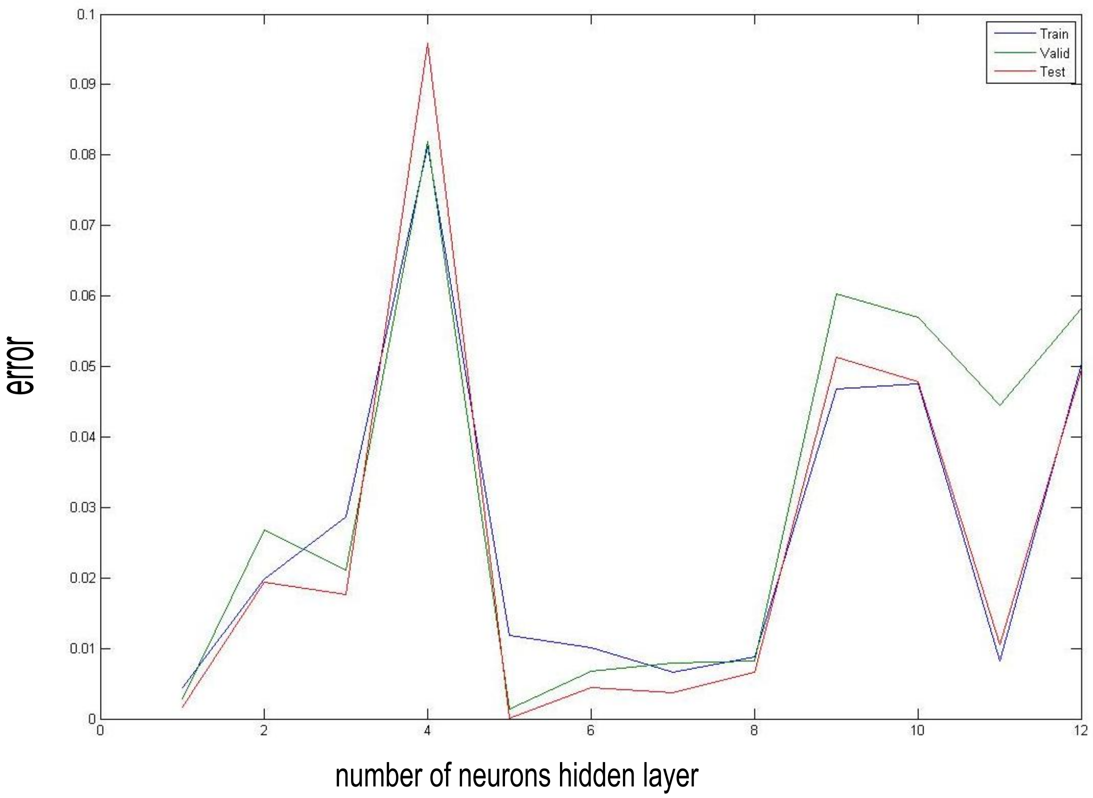


Figure 3-17- Determination of neurons in hidden layer basis on the least error content

Figure 3-18, displays the value of mean proportional error of test datasets.

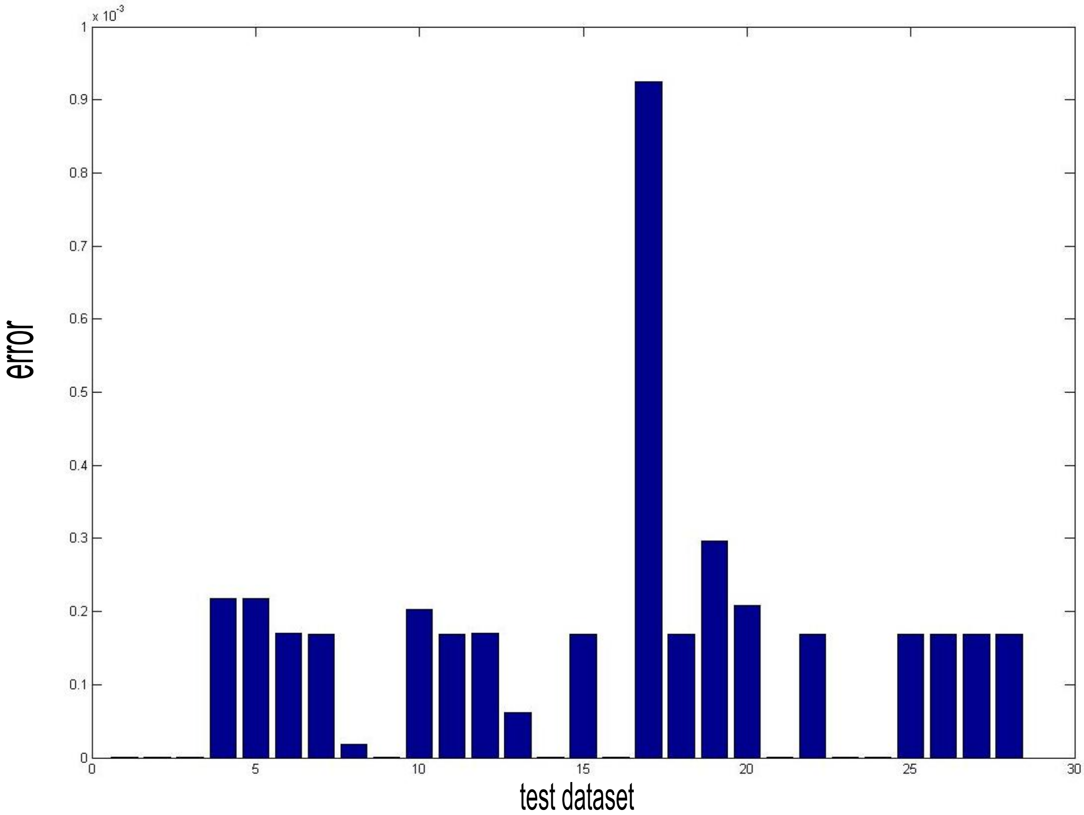


Figure 3-18- Mean proportional error value of test datasets

3-3-5- Modeling the data layers

After training the network, now the data layers can be modeled using the concluded pattern. For this part a program is prepared base on the following steps:

1- Call and read the raster data layers and put them in separate matrix.
2- Save the all data layers in a matrix (kb) with 3 dimensions as: 1875 (number of columns in raster data) * 2399 (number of rows in raster data) * 5 (number of data layers).
3- Put the matrix (kb) in a 2 dimensional matrix (pn), which includes 5 * 4498125 dimensions.
4- Partition the “pn” to 6 parts for easier modeling.
5- Put the modeling result (output), in a matrix (an) with 2 dimensions.
6- Convert the matrix “an” to a matrix with 3 dimensions (A1).
7- Draw the figure A1.

3-3-6- Post-processing

Output values that create the figure “A1” must be rectified to original location. This step can be done using a header file. In header file the georeferencing parameters were stated. The most important referencing parameters for this project are as follow:

```
DataSetType = ERStorage
DataType = Raster
CoordinateSpace Begin
  Datum = "WGS84"
  Projection = "NUTM40"
  CoordinateType = EN
  Units = "METERS"
  RegistrationCoord Begin
    Eastings = 307781
    Northings = 3375987
```

RegistrationCoord End

CellInfo Begin

Xdimension = 60

Ydimension = 60

The georeferenced output can be reclassified for better presentation. This image shows the mineral prospectivity map for desired study area. As it has shown in the legend the higher values imply to more probability of mineralization. Figure 3-18 shows the mineral prospectivity mapping of the study area.

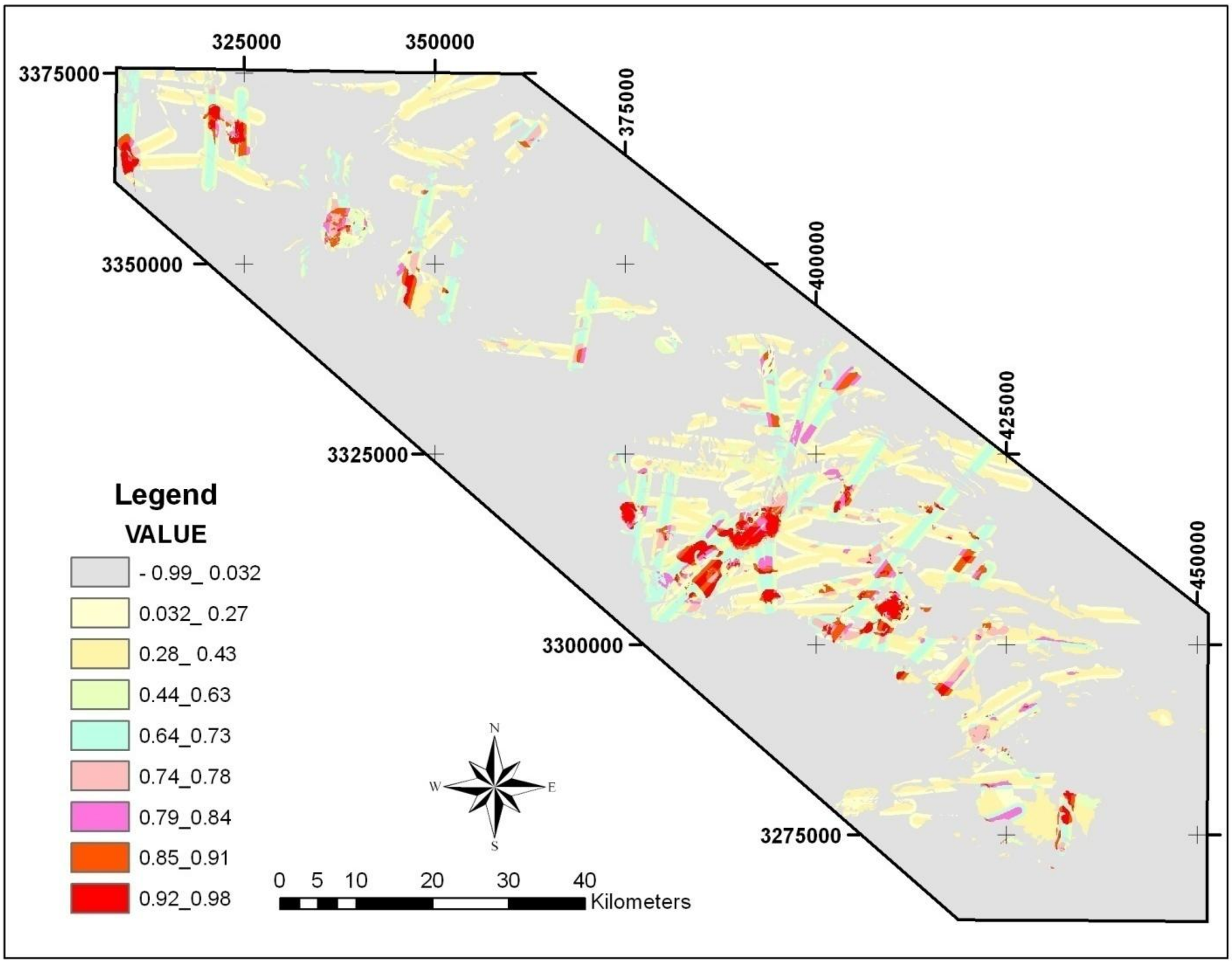


Figure 3-18- Mineral prospectivity mapping of the study area using neural network method

4- Conclusion

Artificial Neural Networks as a data-driven method, that the allocation appropriate weights to layers implement basis on the datasets. Unlike other data-driven methods which need only one set of training points i.e. the site of mineral deposits, ANN requires another set of training points which defines the absence of mineral deposits. The second set of training points was generated using the areas of very low probability of probabilistic map resulted from simple overlaying method. Also the operations for the application of neural networks employed in this study involve multi-class representation of evidential maps. The predictive map generated from multi-class evidential maps showed higher predictive rate with respect to known deposits and indications compared to the predictive map resulted from binary maps.

Characteristics of neural networks that make them particularly suitable for pattern recognition and classification consist of their capability to:

a) Extract underlying patterns in a dataset which may be incomprehensible to humans and standard statistical techniques.
b) Operate at acceptable accuracy when the data are noisy or contain outliers.
c) Flexible and retrained when new data become available.

Numerical Mineral Prospectivity Map can be classified to three (first, second, and third) categorical priorities. Favorable Copper porphyry mineralization area where extracted using artificial neural networks method include 60 targets (figure 4-1), which contain 5.6 percent of study area. The 18 regions locate in old mining area and 42 regions are new area. Statistical characteristics of first class targets are shown in table 4-1. More inquiry implies to high-quality of results using ANN method, because all of positive training datasets locate in first class targets and results don’t miss any training data.

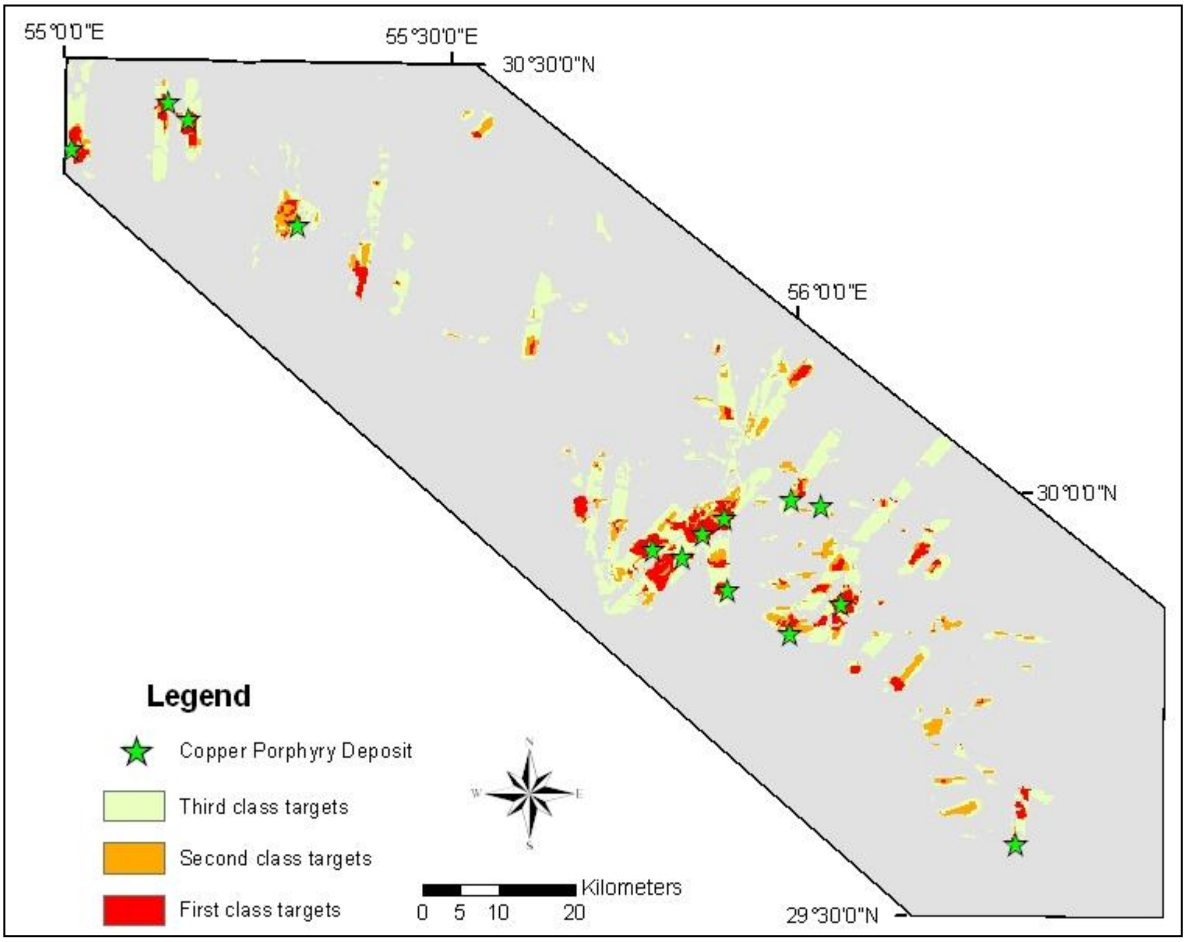


Figure 4-1- Three classes' mineral prospectivity mapping using neural network method

Table 4-1- Statistical characteristics of first class targets using neural network method

Target	Number	Area (km2)
All	**60**	**122**
Old	**18**	**48**
New	**42**	**44**

The results of this research indicate that the neural networks method has several advantages. These contain the capability to: a) react to critical integrations of factors, rather than automatically increasing the prospectivity because of all suitable parameters; b) integrate datasets without the loss of information inherent in existing methods; and c) produce results that are relatively unaffected by unnecessary, false data.

References

Alavi, M., 1994, “Tectonic of the Zagros orogenic belt of Iran: new data and interpretations, Tectonophysics”, 229, pp.211-238.

Alavi, M., 1996, “Tectonostratigraphic synthesis and structural style of the Alborz mountain system in northern Iran”, Journal of Geodynamics, v. 21, p.1–33.

Alavi, M., 2004, “Regional stratigraphy of the Zagros fold-thrust belt of Iran and its proforland evalution”, American Journal of Science, Vol.304, January, pp.1–20.

Amidi, S.M., 1984, “Alkaline character of Eocene volcanism in the middle part of Iran and its geodynamic situation”, Geologischen rundschau, Vol. 73, 917-932.

Bakhshi, B., Khorsandi M., 2007 “Improvement on Concentrator Location Problem in Access Network Design using Growing Neural Network”, CSICC.

Benediktsson J. A., Swain P. H. & Erosy O. K., 1993, “Conjugate-gradient neural networks in classification of multisource and very-high dimensional remote sensing data”, International Journal of Remote Sensing 14, 2883–2903.

Berberian, M., and King, G.C.P., 1981, “Towards a paleogeography and tectonic evolution of Iran”. Canadian Journal of Earth Sciences, v. 18, p. 210–265.

Berberian, F., Muir, I.D., Rankhurst, R.J. and Berberian, M., 1982, “Late Cretaceous and early Miocene Andean-type plutonic activity in northern Makran and Central Iran”, J. Geol. Soc. London, Vol.139, pp.605-614.

Bishop C. M., 1995, “Neural Networks for Pattern Recognition”. Clarendon Press, Oxford.

Brodeur G., Jebrak M., 1992, “Applications of neural network computing to mining exploration in the Southern Abitibi greenstone belt”, In: Program with Abstracts—Geological Association of Canada, Mineralogical Association of Canada, Canadian Geophysical Union Joint Annual Meeting Vol. 17, p. 12. Geological Association of Canada, Acadia University, Wolfville.

Brown W. M., Gedeon T. D., Groves D. I. & Barnes R. G., 2000, “Artificial neural networks: a new method for mineral prospectivity mapping”, Australian Journal of Earth Science, 47, 757-770.

Clare A., Bolt G. & Daneel G. 1997, “The use of neural analysis as an exploration tool”, In: Proceedings of the 12th International Conference on Applied Geologic Sensing Vol. 2, pp. 199–205.

Cucuzza J. & Goode A. D. T., 1998, “Australian mineral exploration research”, In: Berkman D. A. & MacKenzie D. H. eds. Geology of Australian and Papua New Guinean Mineral Deposits, pp. 53–60.

Decateur S. E. 1989, “Application of neural networks to terrain classification”, IEEE INNS International Joint Conference on Neural Networks, Vol. 1, 283–288.

Forster, H., Fesefildt, K and Kursten, M., 1972, “Magmatic and orogenic evolution of the central Iranian voIcanic belt”, 24th International Geology Congress, Section2, pp.198-210.

Glennie, K. W., 2000, “Cretaceous tectonic evolution of Arabia's eastern plate margin: A tale of two oceans, in Middle East models of Jurassic/Cretaccous catbonate systems”, SEPM (Society for Sedimentaly Geology) Spec. Publ. No. 69, pp. 9-20.

Gong P., 1996, “Integrated analysis of spatial data from multiple sources: using evidential reasoning and artificial neural network techniques for geological mapping”, Photogrammetric Engineering and Remote Sensing 62, 513–523.

Grupal Singh Er., Model of neuron, personal online website.

Hassoun M. H., 1995, “Fundamentals of Artificial Neural Networks”, MIT Press, Cambridge.

Hooper, R.J., Baron, I., Hatcher JR., R.D. & Agash, S. 1994, “The development of the southern Tethyan margin in Iran after the break up of Gondwana: implications of the Zagros hydrocarbon province”, Geosciences 4, 72–85.

Hornik K., 1991, “Approximation capabilities of multilayer feedforward networks”, Journal of Australian Geology & Geophysics 17, Neural Networks, 4, 251–257.

Huang, S., Huang, Y., 1991, “Bounds on the number of hidden neurons”, IEEE transactions on neural networks, volume 2, Part 1, p47-55.

Hush, D. R., Horne, B. and Salas, J. M., 1992, “Error surfaces for multilayer”. IEEE Transactions on Systems, Man, and Cybernetics **22**, pp. 1152–1161.

Jafari Rad, Ali Reza, 2009, Modeling of conceptual and empirical geospatial dataset for porphyry copper mineral prospectivity mapping in central Iran, PhD thesis, TU-Clausthal, Germany .

Jankovic, S., 1984, “Meta1logenyoftheAlpinegranitoids in the Tethyan一Eurasian Meta1logenic Belt”, Proceedingofthe 27th Intemational Geologicall Congress, Moscow, August 4-14, v.12, , VNU Science Press, the Netherlands, pp. 247-273.

Jeffrey W. & Rossner R. 1986, “Optimization algorithms: simulated annealing and neural network processing”, Astrophysical Journal 310, 473–481.

Kohonen, T., 1987, “Self-organization and associative memory”, 2nd edition, Berlin: Springer-Verlag.

Masters T., 1993, “Practical Neural Network Recipes”, in C11. Academic Press, San Diego.

The MathWorks, Inc. 2016.

Richard M. D. & Lippmann R. P., 1991, “Neural networks classifications estimate Bayesian a posteriori probabilities”, Neural Computation 3, 461–483.

Ritter N. D. & Hepner G. F., 1990, “Application of an artificial neural network to land-cover classification of thematic mapper imagery”, Computers and Geosciences 16, 873–880.

Ritter N. D., Logan T. L. & Brynat N. A., 1988, “Integration of neural network technologies with geographic information systems”, In: Thomas H. F. ed.

GIS: Integrating Technology and Geoscience Applications (Denver), pp.102–103. National Academy of Science, Washington DC.

Romeo, G., 1994, “Seismic signal detection and classification using artificial neural networks”, Annali di Geofisica 3,343-353.

Rutkowski L., 2008, “Computational Intelligence, methods and techniques, Springer.

Sahandi M. & et al., 2008, “Structural map of Iran”, Geological Survey of Iran.

Samani, B., 1998. “Distribution, setting and metallogenesis of copper deposits in Iran”, Porphyry and gold deposits, global perspective, international conference, Perth, Western Australia, 135-158.

Samani, B. & Ashtari s., 1992, “Evaluation of Volcano – Plutonic – Arc in central Iran, formed by subduction of Mesozoic Tethys oceanic crust”, Geosciences. GSI 1 (4).

Sengor, A. M. C., 1990, “A new model for the Late Paleozoic- Mesozoic tectonic evolution of Iran and implication for Oman Region. In The Geology and Tectonics of the Oman Region”, Geological Society of London, Special Publication 49, pp.797-831.

Shahabpour J., 1982, “Unroofing "fragments" as a reconaissance explora- tion tool in the Central Iranian porphyry copper belt”, Economic Geology, vol. 87, pp.1599-1606.

Shahabpour, J., 2005, “Tectonic evolution of the orogenic belt in the region located between Kerman and Neyriz”. Journal of Asian Earth Sciences, vol.24, pp. 405-417.

Stampfli, G.M. and Borel, G.D., 2002, “A plate tectonic model for the Paleozoic and Mesozoic constrained by dynamic plate boundaries and restored synthetic oceanic ishchrons”, Earth and Planetary Science Letters, 196:17-33.

White H., 1989, “Learning in artificial neural networks: a statistical perspective”. Neural Computation 1, 425–464.

Wong P. M., Taggart I. J. & Gedeon T. D., 1995, “Use of neural network methods to predict porosity and permeability of a petroleum reservoir”, AI Applications 9, 27–37.

Zadeh L. A., 1965, “Fuzzy sets”, IEEE information and control, Vol. 8, 338-353.

Zadeh L.A. & Kacprzyk J., 1992, “Fuzzy Logic for the Management of Uncertainty”, L.A. and J. Kacprzyk (Eds.), John Wiley & Sons, New York.