RESEARCH ARTICLE OPEN ACCESS

A Review of Identification of Minerals Using Remote Sensing and Geographical Information System

Chetan Marlecha *, Vinod Pande**, ***Sangramsing Kayte
*(Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University,Aurangabad, India)
** (Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University,Aurangabad, India)
*** (Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University,Aurangabad, India)

ABSTRACT
Exploration of new mines is vitally important for human life. Geospatial Information Systems (GIS) can be effectively used in the gathering, weighting, analyzing and presenting spatial and attribute information to facilitate the mine exploration process. The success of mine exploration largely depends on: the identification of governing factors, the determination of their impacts and the selection of suitable models to integrate the parameters. Weighting methods are classified into two main groups: data-driven and knowledge-driven. Six weighting methods are identified and scientifically assessed in this study, namely; Ratio Estimation, Analytical Hierarchy Process (AHP), Delphi, Weight of Evidence, Logistic Regression and Artificial Neural Networks (ANN). The first three are examples of knowledge-driven and the last three are classified in the data-driven group. In order to evaluate the methods, the information of 26 copper boreholes are used. Numerical experimentations showed that the artificial neural network used in this study is the most accurate method because it could predict the characteristics of all boreholes correctly. It is shown that knowledge-driven methods are very much affected by the degree of knowledge and the specialization of experts. The results indicated that AHP is the most successful method among knowledge-driven class and could predict the characteristics of 82% of boreholes correctly.
Keywords: Data-driven, knowledge-driven, mineral potential map, artificial neural networks, Rs: remote sensing GIS: geographic information system. Gold Potential Mapping, Analytical Hierarchy Process Delphi,Weight of Evidence, Logistic Regression and Artificial Neural Networks (ANN).

I. INTRODUCTION

Identification of minerals has been done from an ancient's time for making use of minerals for day today life .The mining of minerals have Minerals like water, petroleum, gold, copper and many more. Remote Sensing and Geographic Information System (GIS) have played an active role in mineral exploration by helping in the identification or discovery of new gold deposits in most part of the world such as Spain, Nova Scotia (Canada) and Egypt. Remote sensing and gis now a days has become the part of every sector for their growth. Every field is now taking help of rs and gis for future prospective.Geologists, miners and engineers are dealing with problems related to the analysis and manipulation of geospatial information to explore minerals for many decades. It is a multistage investigation that begins at a small scale maps and progresses to larger ones. At each stage, geological, geochemical and geophysical data are collected, processed and analyzed to produce Mineral Potential Maps (MPM). Even after labor intensive studies on a deposit, predicting the exact location and the amount of minerals under the ground is difficult. Boreholes must be drilled to find out the exact characteristics of the underground deposits. However, drilling is expensive if not impossible. A GIS has the potential for storing, updating, retrieving, displaying, processing, analyzing and integrating various geospatial data. GIS can produce MPMs easily and integrates the results of different investigations such as geological, geophysical and geochemical studies. Using a powerful method for weighting of the information, GIS can provide a better prediction on the potential of mineralization under the ground. The basic pre-requisite for MPM generation is the determination of weights and rating values representing the relative importance of factors and their categories. Determining the relative importance of information is called map layer weighting. In general, each layer of information includes some sub-classes. The importance of sub-classes has to be determined before assigning weights to the layers. This procedure is called calibration and the weights are assigned to the classes are called rating. There are two main methods for weighting the information layers; data-driven and knowledge-driven. In data-driven methods, the importance of data is determined using data itself while in knowledge-driven methods, an expert or a group of experts perform this task. Six methods are implemented in this research to scientifically assess weighting predictors of copper mineralization and producing

MPMs. They are: Analytical Hierarchy Process (AHP), Delphi and Ratio Estimation (RE) (as knowledge driven) and Logistic Regression (LR), Weight Of Evidence (WOE) and Artificial Neural Networks (ANN) (as data-driven methods).

II. KNOWLEDGE-DRIVEN METHODS

The weights and ratings in this category are determined using subjective experts knowledge. Although, it can be implemented in various ways [34], three approaches are used in this investigation: AHP, Delphi and Ratio Estimation.AHP: AHP is a mathematical decision making technique that allows for the rational evaluation of weightings [32]. It determines an optimal solution through the use of simple representation of a hierarchical model. AHP relies on three fundamental assumptions:

- Preferences for different alternatives depend on separate criteria which can be reasoned about independently and given numerical scores.
- The score for a given criteria can be estimated from sub-criteria. That is, the criteria can be arranged in a hierarchy and the score at each level of the hierarchy can be calculated as a weighted sum of the lower level scores.
- At a given level, suitable scores can be calculated from only pair-wise comparisons.

The scores are arranged in a matrix and the weights for each of the compared elements are calculated using various methods such as eigenvector. This gives a weight for each element within a cluster as well as inconsistency ratio[31]. The inconsistency checking can be done through the following relations:

$$\text{II (Inconsistency Index)} = \frac{\lambda_{max} - n}{n-1} \quad \text{........(1)}$$

$$\text{I.R} = \frac{\text{II}}{\text{I.I.R}} \quad \text{........(2)}$$

Where n is the dimension of comparison matrix, λmax is the maximum eigenvalue of the comparison matrix and I.I.R is the inconsistency index of a random matrix with the same dimension as the comparison matrix. Finally, I.R is the inconsistency ratio. If I.R is less than 0.1 the comparisons are consistent, if not they should be compared again [21]. The final weight of each alternative or sub-criteria (in a hierarchy like Fig. 1) is obtained using Equation 3. The sum of calculated weights for each comparison matrix is equal to 1.s

$$\begin{aligned} wA &= w11w1 + w21w \\ wB &= w12w1 + w22w2 \end{aligned} \quad \text{.......... (3)}$$

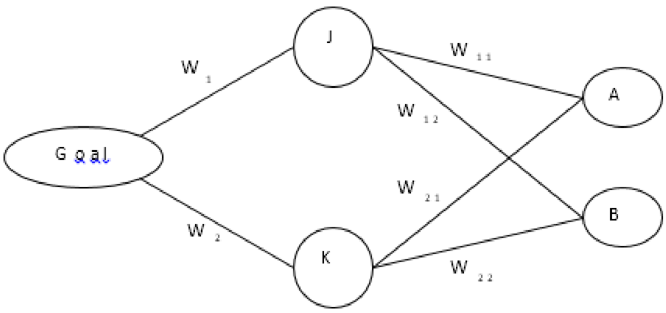

Fig:1 Calculating the Final weights in a Hierarchy of AHP

Delphi: Delphi provides reliable information for weighting [29]. The method gathers knowledge from a group of experts by means of a series of questionnaires and their feedbacks. Anonymity, controlled feedback and statistical response characterize Delphi these characteristics provide distinct advantages over the conventional face-to-face conference as a communication tool. A team is chosen to do the Delphi activities. They invite experts and prepare the questionnaires. The questionnaires are distributed among the participants and after some sessions of conversation participants often converge to unique decisions about each related weight.Ratio Estimation: This approach is categorized in rating methods of weighting First the criteria are ranked from the most important to the least important. Then, the method starts by assigning an arbitrary weight (i.e. 100) to the most important criterion, as well as to the least important attribute. The value for the least important criterion is, then, divided by the score for each criterion: that is, the ratio is equal to wi/w*, where w* is the lowest score and w i is the score for the ith criterion. This ratio expresses the relative desirability of a change from the worst level of that criterion to its best value. This states how much more or less valuable an alternative is than the best, in a ratio sense. This procedure is repeated for the next most important criterion until weights are assigned to all criteria. Finally, the weights are normalized by dividing each weight by the total.

III. DATA-DRIVEN METHODS

Data-driven method reduces the problem of biased or incorrect decisions that knowledge-driven method may have. For minimizing the subjectivity and bias in the weighting process, quantitative methods, namely, statistical analysis, deterministic analysis, probabilistic model and distribution free approaches may be utilized. Data-driven models need samples of results to be executed and evaluated.Weight of evidence (WOE): The WOE is a data-driven and discrete multivariate statistical method that uses conditional probabilities to determine the relative importance of parameters prior and posterior probabilities are the major concepts which are used in this approach to delineate the relative importance of data. If a phenomenon has been assessed in a region, then, it is present in some points and is absent in others. Thus, the probability of occurrence of this phenomenon can be calculated by dividing the number of occurrence samples to the whole assessed samples [30]. This probability is called prior probability. The posterior probability is the conditional probability of existence of the phenomena when the predictor exists. The following equations formulate the basis of WOE [5]:

$$\operatorname{logit}(D \mid B_1 \cap B_2 \cap B_3 ... \cap B_n) = \operatorname{logit}(D) + \sum_{i=1}^{n} W_i^s \quad (4)$$

Where, logit is the natural logarithm, D is an event, B1, B2…Bn are binary maps which are considered as a predictor for D and Wis is the weight which is changed to Wj+ when the predictor Bj is present and also is changed to Wj− when the predictor Bj is absent (Wj+ and Wj− are the positive and negative weights of evidence). If one or more data is not available somewhere, the W is 0 for that area. The contrast C provides a measure of spatial association between a set of occurrence points and an evidence pattern and is derived from:

$$C=W+ -W- \quad (5)$$

If more occurrences occur within a pattern than would be expected by chance, then W+ is positive and W- is negative. In contrary, W+ is negative and W- is positive when fewer points occur within a pattern by chance.Since Equation 4 is derived assuming the conditional independency between predictor maps, it is necessary to evaluate the conditional independency between layers of data before using WOE [9]. The maximum contrast in a large area with a large number of occurrences gives the best measure of spatial correlation For each estimated weight, the variance can be calculated. The sum of variances for two weights is the variance of contrast. Dividing the contrast by its standard deviation, the studentized value can be calculated [1]. The studentized value serves as an approximate test of the spatial association between the occurrence points and the test domain. It is an informal test that C is significantly different than zero, or the contrast is likely to be real. This test is applied when a small area being considered and there is only a small number of occurrence points (In such cases the uncertainty of the weights is large and C is meaningless). Contrast and studentized values are suitable parameters to determine cut off values to produce a binary map from a continuous map [16]. Studentized values can be used to produce uncertainty maps as well. Logistic Regression: Logistic Regression (LR) is a part of statistical models called generalized linear models [20]. LR describes the relationship between the response (dependent) and the linear sum of the predictor (independent) variables. LR can predict a discrete outcome, such as MPM, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these [25]. Generally, the dependent or response variable is dichotomous like presence/absence or success/failure. Logistic regression makes no assumption about the distribution of the independent variables [23]. The presence/absence of copper potential can be transformed into a continuous probability space ranging from 0 to 1[17]. Values close to 1 represent high probability of presence; whereas, values close to 0 represent high probability of absence.

Artificial neural networks (ANN): Artificial neural networks have been used in many branches of science due to their versatile characteristics [13]. An artificial neural network operates by creating connections between many different processing elements, each analogous to a single neuron in a biological brain. Each neuron takes many input signals, then, based on an internal weighting system, it produces a single output signal that is typically sent as input to the other neurons [28]. Ability of learning is one of the most important characteristics of ANN [3]. Based on the type of training, ANNs are categorized into two main classes of supervised and unsupervised networks [6]. The network weights are modified in the training process through a number of learning algorithms based on back propagation learning [7]. The most widely used back propagation algorithms are gradient decent and gradient decent with momentum. A feed forward multilayer network consists of three layers namely; input, output and hidden layers. Each layer in a network contains adequate number of neurons depending on specific applications. The number of neurons in the input layer is equal to the number of data sources and the number of neurons in the output layer is limited by the application and is represented by the number of outputs. The number of hidden layers and the number of neurons in each layer depend on the architecture of network and usually are determined by trial and error [33]. Index overlay: Index overlay is used in this research to integrate various data layers. In index overlay method, each class of maps is given a different score allowing for a flexible weighting system. The table of scores and the map weights can be adjusted to reflect the judgment of experts in the domain of the application under consideration [5]. At any location, the output score S is defined as:

$$\overline{S} = \frac{\sum_{i=1}^{n} S_{ij} W_i}{\sum_{i=1}^{n} W_i} \qquad (6)$$

Where, S is the assigned score to the cell (or polygon), WI is the weight of the ith map and Sij is the weight of jth class from the ith map. When a map is binary Sij will be 0 or 1. The biggest disadvantage of this method probably lies in its linear additive nature. However, as the method is the same for all of the weighting methods it will not have a biased effect.

IV. PRODUCING MPM

Sixteen boreholes are used to build the LR model. Seven boreholes out of 16 are good ones, therefore, the primary probability of being good is 7/16=0.4375. Then, the probability of being a high potential point is calculated. If the probability for each pixel exceeds the value of 0.4375, it is considered as a high potential pixel. Similar method is performed for WOE. Using Equation 4, MPM is estimated. The cut off value in WOE is also considered as 0.4375. Therefore, any cell in MPM surpass the cut-off value is classified in high potential area. The same 16 boreholes are used for training ANN. The output values of ANN do not bear the concept of probability like LR or WOE models. A natural break classifies the MPM into two classes of high potential and low potential [19].The weights extracted from ll models of weightings are entered to Index Overlay model. Then the MPMs are produced by Index Overlay. For the produced MPMs, experts determined the cut-off value for classifying MPMs to two classes of high and low mineral potential. Experts determined 0.4 as a cut-off value for this purpose. Figure 8 represents the final MPMs produced for copper deposit of Ali-Abad using AHP and ANN. In this research, six different weighting procedures viz. Ratio Estimation, Analytical Hierarchy Process (AHP), Delphi, Weight of Evidence, Logistic Regression and Artificial Neural Networks (ANN) were applied for producing MPMs in part of Ali-Abad copper deposit in Iran. A comparative evaluation was also carried out. The ANN approach produced the most accurate map. This may be attributed to the objective approach where weights for factors are determined through ANN connection weight approach. In short, the following observations were made based on this research: Therefore, the integration of different factors in GIS environment using a variety of weighting procedures may serve as one of the key objectives in any MPMs generation. It is recommended that uncertainty in weighting methods to be studied further. It is also recommended that other methods for integrating the weights obtained from data-driven approaches to be tested. Providing alternative solutions to integrate the weights obtained from data-driven and knowledge-driven approaches can be another option for more researches.

V. CONCLUSION

In this research, six different weighting procedures viz. Ratio Estimation, Analytical Hierarchy Process (AHP), Delphi, Weight of Evidence, Logistic Regression and Artificial Neural Networks (ANN) were applied for producing MPMs in part of Ali-Abad copper deposit in Iran. A comparative evaluation was also carried out. The ANN approach produced the most accurate map. This may be attributed to the objective approach where weights for factors are determined through ANN connection weight approach. In short, the following observations were made based on this research:Therefore, the integration of different factors in GIS environment using a variety of weighting procedures may serve as one of the key objectives in any MPMs generation. It is recommended that uncertainty in weighting methods to be studied further. It is also recommended that other methods for integrating the weights obtained from data-driven approaches to be tested. Providing alternative solutions to integrate the weights obtained from data-driven and knowledge-driven approaches can be another option for more researches.

REFERENCES

[1] Agterberg, F. and Q. Cheng, 2002. Conditional Independence Test for Weights-of-Evidence Modeling, Nat. Resour. Res., 11(4): 249-255.

[2] Asadi, H. and M. Hale, 2001. A Predictive GIS Model for Mapping Potential Gold and Base Metal Mineralization in Takab Area, Iran, Computer and Geosci., 27 (8): 901-912.

[3] Beale R. and T. Jackson, 1998. Neural Computing: An Introduction, Institute of Physics Publishng.

[4] Berry, J.K., 2002. Map analysis: Procedures and Applications in GIS Modeling, Basis Press.

[5] Bonham-Carter, G., 1994. Geographic Information Systems for Geoscientists:

Modelling with GIS. Pergamon Press, Oxford, pp: 398.

[6] Brown, W., T. Gedeon, D. Groves and R. Barnes, 2000, Artificial Neural Networks: A new method for mineral prospectivity mapping: Australian J. Earth Sci., 47 (4): 757-770.

[7] Brown, W., D. Groves and T. Gedeon 2003. Use of Fuzzy Membership Input Layers to Combine Subjective Geological Knowledge and Empirical Data in a Neural Network Method for Mineral-

[8] Potential Mapping. Natural Resourc. Res., 12 (3): 183-200.

[9] Calcerrada, R. and S. Luque, 2006. Habitat Quality Assessment using Weight-of-Evidence based GIS modelling: The Case of Picoids Tridactylus as Species Indicator of the Biodiversity Value of the Finnish Forest, Ecological Modelling. 196: 62-76.

[10] Carranza, E. and M. Hale, 2002. Where are porphyry copper deposits spatially localized? A Case Study in Benguet Province, Philippines, 11 (1): 45-59.

[11] Chen, Y., 2004. MRPM: three visual basic programs for mineral resource potential mappin. Comput. and Geosci., 30 (9-10): 969-983.

[12] Ford, A., T. Blenkinsop, 2007. Combining fractal analysis of mineral deposit clustering with weights of evidence to evaluate patterns of mineralization: Application to copper deposits of the Mount Isa Inlier, NW Queensland, Australia. Ore Geology Reviews. Available on line 7 March 2007.

[13] Fowles, J., 1978. Handbook of Futures Research. Westport CT, Greenwood Press. Connecticut.

[14] Graupe, D., 2007. Principles of Artificial Neural Networks, Advanced Series in Circuits and Systems. 2nd Edn. World Scientific Publishing Company, pp: 250.

[15] Hanne, T., 2000. Intelligent Strategies for Meta Multiple Criteria Decision Making, Springer, pp: 236.

[16] Harris, J., Wilkinson, L. and Grunsky, E., 2000.

[17] Effective use and interpretation of lithogeochemical data in regional mineral exploration programs: Application of Geographic Information Systems (GIS) technology. Ore Geolo. Rev., 16 (3-4): 107-143.

[18] Harris, J., L. Wilkinson, K. Heather, S. Fumerton, M. Bernier, J. Ayer and R. Dahn, 2001. Application of GIS processing techniques for producing mineral prospectivity maps - A case study: mesothermal Au in the Swayze Greenstone Belt, Ontario, Canada. Nat. Resour. Res., 10 (2): 91-124.

[19] Hosm r D. and S. Lem show, 2000. Applied Logistic Regression, 2nd Edn., John Wiley and Sons Inc.

[20] Huang, S., 1997. Application of optimum stratified sampling and separate ratio estimation in stochastic copper loss of transmission system evaluation IEE Proceedings-Generation, Transmission and Distribution, 144 (2): 207-212.

[21] Kanungo, D., M. Arora, S. Sarkar and R. Gupta, 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas, Engineering Geol. 85: 347-366.

[22] Kleinbaum, D. and M. Klein, 2005. Logistic Regression. 2nd Edition. Springer, pp: 536.

[23] Kolat, C., V. Doyuran, C. Ayday and L. Süzen, 2006. Preparation of a geotechnical microzonation model using Geographical Information Systems based on Multicriteria Decision Analysis. Engineering Geology. 87: 241-255.

[24] Ligas, P. and M. Palomba, 2006. An integrated application of geological-geophysical methodologies as a cost-efficient tool in improving the estimation of clay deposit potential: Case study from South-Central Sardinia (Italy). Ore Geology Reviews. 29 (2): 162-175.

[25] Long, J. and Freese, J., 2005. Regression Models for Categorical Dependent Variables using Stata, Stata Press, pp: 528.

[26] Malczewski, J., 1999. GIS and Multicriteria Decision Analysis. John Wiley and Sons INC. pp: 392.

[27] Menard, S., 2001. Applied Logistic Regression Analysis, 2nd Edn. Sage Publications Inc. pp: 128.

[28] Moon, C., M. Whateley and A. Evans, 2006. Introduction to Mineral Exploration, 2nd Edn. Blackwell Publishing, pp: 496.

[29] Porwal, A., E. Carranza and M. Hale, 2003. Knowledge-driven and data-driven fuzzy models for predictive mineral potential mapping: Natural Resources Research, 12 (1): 1-25.

[30] Porwal, A., E. Carranza and M. Hale, 2003a. Artificial Neural Networks for Mineral-Potential Mapping: A Case Study from Aravalli Province. Natural Resources Research, 12 (3): 155-171.

[31] Rowe, G. and G. Wright, 2001. Expert opinions in forecasting: The role of the Delphi technique, in J. S. Armstrong (ed.), Principles of Forecasting. Boston: Kluwer Academic Publishers, pp: 125-144.
[32] Roy, R., D. Cassard, P. Cobbold, E. Rosse lo, M. Billa, L. Bailly and A. Lips, 2006. Predictive mapping for copper-gold magmatic-hydrothermal systems in NW Argentina: Use of a regional-scale GIS, application of an expert-guided data-driven approach and comparison with results from a continental-scale GIS. Ore Geology Reviews, 29 (3-4): 260-286.
[33] Saaty, T., 1980. The Analytic Hierarchy Process. NY, McGraw Hill.
[34] Saaty, T., 2001. Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World. RWS Publications Pittsburg. pp: 323.
[35] Samanta, B., S. Bandopadhyay nd R. Ganguli, 2006. Comparative Evaluation of Neural Network Learning Algorithms for Ore Grade Estimation. Mathematical Geology, 38 (2): 175-197.
[36] Tsiporkova, E. and V. Boeva, 2006. Multi-step ranking of alternatives in a multi-criteria and multiexpert decision making environment. Information Sciences, 176 (18): 2673-2697.
[37] Zhou, W., G. Chen, H. Li, H. Luo and S. Huang,
[38] 2007. GIS application in mineral resource analysisA case study of offshore marine placer gold at Nome, Alaska. Computers and Geosciences, 33 (6): 773-788