ESDAS: Explorative Spatial Data Analysis Scale to predict spatial structure of landscape

M.Gangappa Dr.C.Kiran Mai Dr.P.Sammulal

Abstract: Cartographic scale definition of the hydrographic features is critical to display the geospatial data as computer aided maps. Discovering the landscape type that provides geographic context is significant to define cartographic scale. Extensive coverage of the spatial gradients are generally used to distinguish the landscape type. The variation or change over from one type to other type landscape are with limited outliers and not instantaneous. Hence the classification of the landscape types is challenging. An exploratory spatial data analysis scale to learn and predict the landscape types in spatial data is described. This research defined a meta-heuristic scale to perform deep learning that extracts patterns from labeled spatial data of landscape types and further classifies the given unlabeled spatial data in to divergent landscape types. Cross validation and misclassification rate analysis are used to evaluate the proposed model. The paper explored the proposed explorative scale and its impact on the selection of landscape regions. The experiments evinced the scalability and robustness of the proposed explorative scale sensibility to recognize landscape diversity.

Keywords: *ESDAS, Hydrography, physiographic regions, landscape, geospatial data, Spatial Structures, Machine Learning*

1 INTRODUCTION

The geographic context helps to understand the geospatial variations. The geographic context can be obtained from the characteristics of the structure and types of Land [1] [2] [3]. In other dimension the type detection of a landscape helps to define cartographic scale to represent hydrographic features. The heterogeneity of features those representing a specific type of landscape is the most critical factor to generalize the cartographies of the hydrographic features [4].

The terrain coarseness, wetness, and land covering structure are few among the significant factors those often uses to characterize the physiographic regions (also known as landscape). The methods that are often uses to define the landscape can distinguish by their processes such as manual definition and computer aided automatic region detection approaches.

- M.Gangappa: AssociatProfessor, Dept of CSE, VNR VignanaJyothi Institute of Engineering and Technology, Hyderabad, Telangana,India. Email: gangappa_m@vnrvjiet.in
- Dr.C.Kiran Mai: Professor in Dept of CSE, VNR VignanaJyothi Institute of Engineering and Technology, Hyderabad,Telangana,India. Email: ckiranmai@vnrvjiet.in
- Dr.P.Sammulal: Assistant Professor in Dept of CSE, JNTUH College of Engineering Nachupally (Kondagattu), Kodimyalmandal, Karimnagar Dist. Telangana. 505 501 India. Email: sammulalporika@gmail.com

This paper draws upon a maximum likelihood classification furnishing physiographic categorization of the given spatial data based on the features most frequently observed and specific to the types of landscape observed in labeled data given for training. Hence, unlike other traditional approaches, the dependency on specific predetermined features is nullified in this proposed model.

The goal of this research is to define an explorative scale to refine the existing classification strategies, such that a classifier can distinguish the classes according to the features found as significant in given training data to furnish the optimal classification of the landscape (types).The types of landscapes adapted to explore the proposed approach are

- The spatial features of Soil structure (wetness, dryness, type)
- Spatial features of Forestry structure (plants, forests, groves and arboretums)
- Valley Landscape Structure
- Water resource structure

The experimental study targeted to estimate the prediction accuracy by cross validation and misclassification scope.

The rest of the paper describes the related work in section 2, proposed Explorative Spatial Data Analysis Scale in section3 that followed by experimental study and performance analysis in section 4 and section 5 concludes the manuscript.

2 RELATED WORK

The recent literature substantially evincing the role of SVM (support vector machines) to model the geospatial data with data validation accuracy [5] [6] [7]. Pozdnoukhov et al., [5] proposed an SVM based modeling of spatial data in order to define and redefine the monitoring networks. This model aimed to discover the prospective positions emerging measurement points. The model adapted the active learning strategy hence the prospective positions will be embedded to knowledge dynamically. The curse of dimensionality is often found in high dimensional space with limited data [8] [9].In order to this the appropriate methods and tools should be used to handle the curse dimensionality [10]. Hence the SVM, which insensible to the dimensionality of the space are superior [11]. Hence the existing models adapted the SVM in order to perform classification [5] [6] [7].

The majority of contributions found in contemporary literature related to prediction of physiographic regions by classifying the geospatial data are based on the machine learning algorithms [12] [13] [14][15][16][17]. The commonality of these models can be found at training phase. AT training phase of these methods, the data with predefined class labels are used to train the learning algorithms like Random Forest, Support Vector Machines and Artificial Neural Networks. The machine learning algorithms Random Forest (RF) and Support Vector Machines (SVM)are found to be optimal that compared to other such as Artificial Neural Networks (ANN).This is due to the insensitivity of RF and SVM towards dimensionality of the search space. Another commonality of all these studies is the use of predetermined features such as mean elevation, deviation of the elevation, density of bed-rock, and surface water occupancy as inland area, overflow, and mean slope. The overall assessment of these existing models indicating that the prediction accuracy depends on the given input spatial data for training. The prediction accuracy at its peak, if the labeled data given for training is evincing the significant contribution of features considered, else prediction accuracy is utterly low. Due to the extensive span of physio graphic regions and gradual variations between types of landscape and considerable absence of the features adopted for training, the prediction accuracy substantially varies from one to other input geospatial data. Another constraint of these existing models is that the features such as spatial domain and the physiographic region expansiveness are not considered [18] [19]. Unlike many of other domains the presence of expected features are sporadic in geospatial data, hence we argue that the selection features should be dynamic and specific to the given labeled data.

3 EXPLORATIVE SPATIAL DATA ANALYSIS SCALE TO PREDICT SPATIAL STRUCTURE OF LANDSCAPE

The ESDAS is an explorative scale that extracts patterns from the given labeled spatial structure data of landscape types Soil structure, forestry structure, Valley Landscape Structure and Water Resource Structure, and further defines a heuristic scale based on the correlation and variance observed between these patterns. In order to this the given spatial structures are partitioned into their respective categories of landscape types adapted.

The spatial attributes involved in each spatial structure are considered as features of the respective category of landscape. Since the spatial structure contains dense number of spatial attributes and majority of them may be insignificant to respective landscape category of the spatial structure. Henceforth, the feature optimization process (see sec 3.1) will be carried out to eliminate these insignificant features. The spatial attribute value range will be discretized further to compare two spatial attributes through equality by approximation (see sec 3.2). Afterwards the confidence of each feature towards all categories of spatial structure data will be assessed (see sec 3.3) that follows the assessment of each spatial structure confidence against the features of all categories (see sec 3.4). Further the confidence obtained for each feature and spatial structure of respective category will be used as input to define the explorative spatial data analysis scales to estimate the scope of Soil structure, the forestry structure, Valley Landscape Structure and Water resource structure.

3.1 Feature Optimization

For each spatial structure context, the spatial structure dataset $S_i = \{s(i)_1, s(i)_2, ...s(i)_{|S_i|}\}$ of size $|S_i|$ will be considered for training towards defining explorative spatial data analysis scale. Each spatial structure is represented by values obtained for sequence of spatial attributes selected from respective spatial structure context. This description binds to all datasets of spatial structures representing Soil structures, Forestry structure, Valley Landscape Structure and water resource structure.

Let $S_n = \{s(n)_1, s(n)_2, ..., s(n)_{|S_n|}\}$ be the set of spatial structures collected from all categories of the spatial structures except D_i. The sets $A_i = \{a(i)_1, a(i)_2, ..., a(i)_{|A_i|}\}$ and $A_n = \{a(n)_1, a(n)_2, ..., a(n)_{|A_n|}\}$ are sequence of spatial attributes representing the features of spatial structures of S_i and S_n respectively. The feature value (spatial attribute value) set $V(i)_j = \{v(ij)_1, v(ij)_2, ...v(ij)_{|V(i)_j|}\}$ be the set of spatial attribute values (feature values) observed for feature $a(i)_j$ of spatial structures represented by S_i. Similarly the feature value set $V(n)_j = \{v(nj)_1, v(nj)_2, ...v(nj)_{|V(n)_j|}\}$ be the set of spatial attribute values observed for spatial attribute $a(n)_j$ of spatial structures represented by S_n. Further discussion refers the feature value (spatial attribute value) as spatial element.

Since the spatial structure is the combination of numerous count of spatial attributes, the size of feature set can lead to process complexity. In order to overcome the process complexity, the insignificant spatial attributes should be identified and discarded. The spatial attribute $a(i)_j$ of A_i is said to be insignificant feature, if spatial elements $V(i)_j$ of $a(i)_j$ are almost similar to the spatial elements $V(n)_j$ of feature $a(n)_j$ of A_n. Hence to identify the insignificant features, we adopt hamming distance that applied on values of each feature as vector from specific spatial structure and all other special structures. The hamming distance with 0 or less than the given threshold indicates that the respective feature is insignificant. The process of hamming distance is explored below:

3.1.1 Hamming Distance

The value of Hamming Distance obtained here is to denote the difference between spatial elements of same feature from specific spatial structure data to all other categories of spatial structure data. This is one of the significant strategy to assess the difference between two spatial elements in coding theory.

The hamming distance between given vectors $CX = \{cx_1, cx_2, ..., cx_n\}$ and $CY = \{cy_1, cy_2, ..., cy_m\}$ of size n and m respectively will be measured as follows:

Let $CZ \leftarrow \phi$ // is a vector of size 0

$foreach\ \{i \exists i = 1,2,3,.....\max(n.m)\}$ Begin

$if\ (\{cx_i \exists cx_i \in CX\} - \{cy_i \exists cy_i \in CY\}) \equiv 0\ then$

3.1.2 $CZ \leftarrow \{cx_i \exists cx_i \in CX\} - \{cy_i \exists cy_i \in CY\}$

Else

$CZ \leftarrow 1$

End

$$hd_{CX \leftrightarrow CY} = \sum_{j=1}^{|CZ|} CZ\{i\}$$

// $hd_{CX \leftrightarrow CY}$ is the hamming distance between CX and CY, $CZ\{i\}$ is the i^{th} element of the vector CZ and $|CZ|$ is the size of the vector CZ

3.2 Spatial Element and spatial structure Confidence Assessment

The spatial elements of optimal features and the spatial structures of respective data set will be used further to assess the spatial element and spatial structure confidence.

In order to this, initially the spatial element pairs will be defined such that each pair contains two spatial elements (feature values) and each feature representing different feature of the same dataset. Then we assess the associability support of each spatial attribute pair. The associability support can be described as the ratio of spatial structures contains that pair against the total number of spatial structures in respective dataset. The process of assessing associability support of each spatial attribute pair is described in following section (see sec 3.2.1).

3.3 Assessing spatial attribute pair correlation

Let P_i be the set and contains all possible unique spatial element pairs from respective dataset S_i. The possible unique spatial element pairs will found as follows:

For each spatial structure $s(i)_j$ of respective dataset S_i, find all possible unique pairs of spatial attributes and add to P_i.

Then correlation of each pair $\{p_j \exists p_j \in P_i\}$ as follows.

Let $\{v_k \exists v_k \in p_j\}$ and $\{v_l \exists v_l \in p_j\}$ be the two spatial attribute values paired as $\{p_j \exists p_j \in P_i\}$, then the correlation $corr(p_j)$ of the pair p_j is

$$corr(p_i) = \frac{\sum_{m=1}^{|S_i|} \{1 \exists \{v_k, v_l\} \subseteq s(i)_m\}}{|S_i|}$$

//The ratio of number of spatial structures contain both spatial attributes against total number of spatial attributes

The correlation of each pair of spatial attributes found in spatial structures of each respective spatial structure data set of Soil structure, forestry structure, Valley Landscape Structure, water source structure should be estimated using the process explored in sec 3.2.1.

3.3.1 Assessing Spatial element and Spatial structure Confidence

In order to assess the confidence of spatial attributes and spatial structures of respective dataset S_i, a mutual relation graph will be formed between spatial structures and spatial attributes of respective S_i. There will be an edge between a spatial attribute and spatial structure if and only if the selected spatial attribute exists in that spatial structure. Then each edge between spatial attribute and spatial structure is weighted as follows.

$$\overset{|V(i)|}{\underset{j=1}{\forall}}\{v_j \exists v_j \in V(i)\} \text{ Begin}$$

$$\overset{|S_i|}{\underset{k=1}{\forall}}\{s(i)_k \exists v_j \in s(i)_k\} \text{ Begin}$$

$$w_{v_j} = 0$$

$$\overset{|s(i)_k|}{\underset{m=1}{\forall}}\left\{v_m \exists v_m \in s(i)_k \wedge v_j \neq v_m\right\} \text{ Begin}$$

$$p_m = \{v_j, v_m\}$$

$$w(v_j) += corr(p_m)$$

End

$$w_{v_j \leftrightarrow s(i)_k} = \frac{w(v_j)}{|s(i)_k| - 1}$$

End

End

The weights obtained for edges between spatial attributes and spatial structures in mutual graph are further used to assess the spatial attribute and spatial structure confidence towards respective Soil structure, forestry structure, Valley Landscape Structure, Valley Landscape Structure and Water Resource structures.

Further we measure the each feature confidence towards spatial structure dataset S_i as follow

$$\overset{|v(i)|}{\underset{j=1}{\forall}}\{v_j \exists V(i) \ni v_j\} \text{ Begin}$$

$$c_{v_j \Rightarrow S_i} = \sum_{k=1}^{|S_i|}\{w(v_j) \exists s(i)_k \ni v_j \wedge S_i \ni s(i)_k\}$$

//aggregating the weight of spatial attribute v_j towards each spatial structure $s(i)_k$ of respective dataset S_i and the same is considered as the respective spatial element confidence towards dataset S_i

End

Similarly each respective spatial structure confidence towards spatial structure dataset S_i is measured as follows

$$\overset{|S_i|}{\underset{j=1}{\forall}}\{s(i)_j \exists S_i \ni s(i)_j\} \text{ Begin}$$

$$c_{s(i)_j \Rightarrow S_i} = \sum_{k=1}^{|V(i)|}\{w(v_k) \otimes c_{v_k \Rightarrow S_i} \exists s(i)_j \ni v_k \wedge S_i \ni s(i)_j\}$$

// the sum of product of each spatial attribute weight and the respective spatial element confidence, such that the spatial attribute exists in selective spatial structure is the confidence of that spatial structure

End

The confidence of spatial attributes and spatial structures of each respective spatial structure data set of soil structure, forestry structure, valley landscape structure and Water Resource Structure should be estimated using the process explored in sec 3.2.2.

3.4 Defining explorative spatial data analysis scales to divergent landscape structures adapted

Further the confidence of spatial structures of datasets S_{sl}, S_{fs}, S_{vs} and S_{ws}

$$m_{sl} = \frac{\sum_{i=1}^{|S_{sl}|}\{c_{s(sl)_i \Rightarrow S_{sl}} \exists S_{sl} \ni s(sl)_i\}}{|S_{sl}|}$$ //Aggregate mean of the respective spatial structures confidence of Soil structure dataset S_{sl}

In order to identify the lower and upper bounds of m_{sl}, the mean absolute distance of S_{sl} is assessed as follows

$$m_{sl}^{ae} = \frac{\sum_{i=1}^{|S_{sl}|}\sqrt{\left(m_{sl} - c_{s(sl)_i \Rightarrow S_{sl}}\right)^2}}{|S_{sl}|}$$

Then the lower and upper bounds of m_{sl} is assessed as

$ml_{sl} = m_{sl} - m_{sl}^{ae}$ // lower bound of m_{sl}

$mu_{sl} = m_{sl} + m_{sl}^{ae}$ // upper bound of m_{sl}

Similarly explorative spatial data analysis scales for S_{fs} (forestry structure), S_{vs} (Valley Landscape Structure) and S_{ws} (Water Resource Structure)

$$m_{fs} = \frac{\sum_{i=1}^{|S_{fs}|}\{c_{s(fs)_i \Rightarrow S_{fs}} \exists S_{fs} \ni s(fs)_i\}}{|S_{fs}|}$$ //Aggregate mean of the respective spatial structures confidence of Forestry Structure dataset S_{fs}

The mean absolute distance of S_{fs} is

$$m_{fs}^{ae} = \frac{\sum_{i=1}^{|S_{fs}|}\sqrt{\left(m_{fs} - c_{s(fs)_i \Rightarrow S_{fs}}\right)^2}}{|S_{fs}|}$$

Then the lower and upper bounds of m_{fs} is assessed as

$ml_{fs} = m_{fs} - m_{fs}^{ae}$ // lower bound of m_{fs}

$mu_{fs} = m_{fs} + m_{fs}^{ae}$ // upper bound of m_{fs}

$$m_{vs} = \frac{\sum_{i=1}^{|S_{vs}|}\{c_{s(vs)_i \Rightarrow S_{vs}} \exists S_{vs} \ni s(vs)_i\}}{|S_{vs}|}$$ //Aggregate mean of the respective spatial structures confidence of Valley Landscape Structure dataset S_{vs}

The mean absolute distance of S_{vs} is

$$m_{vs}^{ae} = \frac{\sum_{i=1}^{|S_{vs}|}\sqrt{\left(m_{vs} - c_{s(vs)_i \Rightarrow S_{vs}}\right)^2}}{|S_{vs}|}$$

Then the lower and upper bounds of m_{vs} is assessed as

$ml_{vs} = m_{vs} - m_{vs}^{ae}$ // lower bound of m_{vs}

$mu_{vs} = m_{vs} + m_{vs}^{ae}$ // upper bound of m_{vs}

$$m_{ws} = \frac{\sum_{i=1}^{|S_{ws}|}\{c_{s(ws)_i \Rightarrow S_{ws}} \exists S_{ws} \ni s(ws)_i\}}{|S_{ws}|}$$ //Aggregate mean of the respective spatial structures confidence of Water Resource Structure dataset S_{ws}

The mean absolute distance of S_{ws} is

$$m_{ws}^{ae} = \frac{\sum_{i=1}^{|S_{ws}|}\sqrt{\left(m_{ws} - c_{s(ws)_i \Rightarrow S_{ws}}\right)^2}}{|S_{ws}|}$$

Then the lower and upper bounds of m_{ws} is assessed as

$ml_{ws} = m_{ws} - m_{ws}^{ae}$ // lower bound of m_{ws}

$mu_{ws} = m_{ws} + m_{ws}^{ae}$ // upper bound of m_{ws}

3.5 Predicting the state of spatial structure

The explorative spatial data analysis scales devised (see section 3.3) will be used further to assess the spatial structure type (soil, forestry, valley or water structure) of a given spatial structure e. The confidence of given spatial structure

$$c_{s \Rightarrow S_{sl}} = \frac{\sum_{i=1}^{|V(S_{sl})|}\left\{c_{v_i \Rightarrow sl} \otimes w(v_i) \exists v_i \in V(S_{sl}) \wedge s \ni v_i\right\}}{\sum_{j=1}^{|V(S_{sl})|}\left\{c_{v_j \Rightarrow sl} \otimes w(v_j) \exists v_j \in V(S_{sl})\right\}}$$

// the aggregate of product of each spatial element confidence and weight, which divides by the aggregate of confidence of all spatial elements exist in $V(S_{sl})$.

Further the confidence of S towards S_{fs}, S_{vs} and D_{ws} assessed as:

$$c_{s \Rightarrow S_{fs}} = \frac{\sum_{i=1}^{|V(S_{fs})|}\left\{c_{v_i \Rightarrow fs} \otimes w(v_i) \exists v_i \in V(S_{fs}) \wedge s \ni v_i\right\}}{\sum_{j=1}^{|V(S_{fs})|}\left\{c_{v_j \Rightarrow fs} \otimes w(v_j) \exists v_j \in V(S_{fs})\right\}}$$

// the aggregate of product of each spatial element confidence and weight, which divides by the aggregate of confidence of all spatial elements exists in $V(S_{fs})$.

$$c_{s \Rightarrow S_{vs}} = \frac{\sum_{i=1}^{|V(S_{vs})|}\left\{c_{v_i \Rightarrow vs} \otimes w(v_i) \exists v_i \in V(S_{vs}) \wedge s \ni v_i\right\}}{\sum_{j=1}^{|V(S_{vs})|}\left\{c_{v_j \Rightarrow vs} \otimes w(v_j) \exists v_j \in V(S_{vs})\right\}}$$

// the aggregate of product of each spatial element confidence and weight of that exists in $V(S_{vs})$ and e, which divides by the aggregate of confidence of all spatial attributes exists in $V(S_{vs})$.

$$c_{s \Rightarrow S_{ws}} = \frac{\sum_{i=1}^{|V(S_{ws})|}\left\{c_{v_i \Rightarrow ws} \otimes w(v_i) \exists v_i \in V(S_{ws}) \wedge s \ni v_i\right\}}{\sum_{j=1}^{|V(S_{ws})|}\left\{c_{v_j \Rightarrow ws} \otimes w(v_j) \exists v_j \in V(S_{ws})\right\}}$$

// the aggregate of product of each spatial element confidence and weight of that exists in $V(S_{ws})$ and e, which divides by the aggregate of confidence of all spatial attributes exists in $V(S_{ws})$.

Then these confidence values of spatial structure e with respect to S_{sl}, S_{fs}, S_{vs} and S_{ws} will be used to estimate the given expression state is Water Resource Structure, Soil structure, Forestry structure or Valley Landscape Structure according to the following conditions.

$\begin{pmatrix}(c_{s \Rightarrow sl} \geq mu_{sl}) \wedge (c_{e \Rightarrow fs} \leq m_{fs}) \wedge \\ (c_{e \Rightarrow vs} \leq m_{vs}) \wedge (c_{e \Rightarrow ws} \leq m_{ws})\end{pmatrix} \Vert$ $\begin{pmatrix}(c_{s \Rightarrow sl} \geq m_{sl}) \wedge (c_{e \Rightarrow fs} \leq ml_{fs}) \wedge \\ (c_{e \Rightarrow vs} \leq ml_{vs}) \wedge (c_{e \Rightarrow ws} \leq ml_{ws})\end{pmatrix}$	Soil structure
$\begin{pmatrix}(c_{s \Rightarrow fs} \geq mu_{fs}) \wedge (c_{e \Rightarrow sl} \leq m_{sl}) \wedge \\ (c_{e \Rightarrow vs} \leq m_{vs}) \wedge (c_{e \Rightarrow ws} < m_{ws})\end{pmatrix} \Vert$ $\begin{pmatrix}(c_{s \Rightarrow fs} \geq m_{fs}) \wedge (c_{e \Rightarrow sl} \leq ml_{sl}) \wedge \\ (c_{e \Rightarrow vs} \leq ml_{vs}) \wedge (c_{e \Rightarrow ws} < ml_{ws})\end{pmatrix}$	forestry structure

$\begin{pmatrix}(c_{s\Rightarrow vs} \ge mu_{vs}) \wedge (c_{e\Rightarrow sl} \le m_{sl}) \wedge \\ (c_{e\Rightarrow fs} \le m_{fs}) \wedge (c_{e\Rightarrow ws} \le m_{ws})\end{pmatrix} \| \begin{pmatrix}(c_{s\Rightarrow vs} \ge m_{vs}) \wedge (c_{e\Rightarrow sl} \le ml_{sl}) \wedge \\ (c_{e\Rightarrow fs} \le ml_{fs}) \wedge (c_{e\Rightarrow ws} \le ml_{ws})\end{pmatrix}$	Valleystructure
$\begin{pmatrix}(c_{s\Rightarrow ws} \ge mu_{ws}) \wedge (c_{e\Rightarrow sl} \le m_{sl}) \wedge \\ (c_{e\Rightarrow fs} \le m_{fs}) \wedge (c_{e\Rightarrow vs} \le m_{vs})\end{pmatrix} \| \begin{pmatrix}(c_{s\Rightarrow ws} \ge m_{ws}) \wedge (c_{e\Rightarrow sl} \le ml_{sl}) \wedge \\ (c_{e\Rightarrow fs} \le ml_{fs}) \wedge (c_{e\Rightarrow vs} \le ml_{vs})\end{pmatrix}$	Water resource structure

4 EXPERIMENTAL STUDY

The experimental study was carried out on a set of spatial structures taken from multiple benchmark datasets [19]. The number of spatial structures used are 1114, which are the combination of Soil structure (286 records), Forestry structure (275 records), Valley Landscape Structure (277 records) and Water Resource Structure condition (276 records). The spatial structures of respective category are considered as separate datasets labeled as S_{sl}, S_{fs}, S_{vs} and S_{ws}. Each dataset partitioned into test and training sets. The 75% of spatial structures of each dataset are considered as training set and rest 25% of spatial structures considered as test set. In order to estimate the specificity (recognizing untrained spatial structure type prediction) of the proposed model, the significant percent (15%) of uncategorized spatial structures included into test data.

The explorative spatial data analysis scales obtained from the given training set were explored in table 1.

m_{sl}	0.582474187
m_{sl}^{ae}	0.095593654
ml_{sl}	0.486880533
mu_{sl}	0.678067841
m_{fs}	0.615957277
m_{fs}^{ae}	0.103864099
ml_{fs}	0.512093178
mu_{fs}	0.719821376
m_{vs}	0.646638853
m_{vs}^{ae}	0.099722167
ml_{vs}	0.546916686
mu_{vs}	0.74636102
m_{ws}	0.631593026
m_{ws}^{ae}	0.068999373
ml_{ws}	0.562593653
mu_{ws}	0.700592398

Table 1: The explorative spatial data analysis scales obtained from training data

True Positives	265
True Negatives	40
False Positives(records predicted wrongly)	12
False Negative (records are not in either of the 4 landscape types)	5
Categorized (Soil, forestry, Valley Landscape and water resource) spatial structure Prediction Value (positive prediction value, PPV)	0.9566787
uncategorized spatial structure Prediction Value (Negative Prediction value, NPV)	0.888888889
Detection Accuracy	0.947204969
Categorized (Soil, forestry, Valley Landscape and water resource) spatial structure prediction Rate (True Positive Rate or sensitivity)	0.946428571
Uncategorized spatial structure Prediction rate (True Negative Rate or specificity)	0.952380952

Table 2: The prediction statistics of the ESDAS

The 322(soil structure: 72, forestry structure: 69, valley landscape structure: 70, water resource: 69 and uncategorized:42) spatial structures were used to assess the prediction accuracy of the proposed ESDAS. The ESDAS assessed the given input spatial structures such that 265 spatial structures are true positives (the detection of soil, forestry, valley landscape and water resource structures are true), 12 spatial structures are false positive (2 uncategorized spatial structures predicted as valley structure and 10 are falsely detected as soil structure, forestry structure, valley landscape structure or water resource structure), 40 spatial structures are true negatives (detecting spatial structures as uncategorized spatial structure is true) and 5 spatial structures are false negative (detecting spatial structures as uncategorized structure is false).

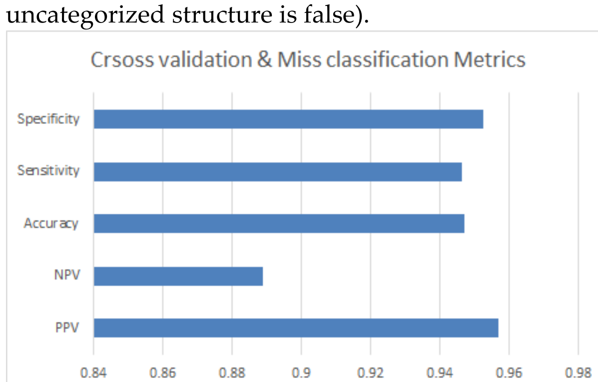

Fig 1: Statistical Metrics and their values observed

Hence the positive prediction value (landscape structure prediction value) is 0.96, negative prediction value (uncategorized structure prediction value) is 0.89, the landscape structure detection rate (also known as sensitivity) is 0.95, the uncategorized structure detection rate (also known as specificity) is 0.95 and the overall success rate (also known as accuracy, which is the ratio between true prediction of all types of spatial structures and all given number of spatial structures) is 0.95. These statistics indicating that the ESDAS is significant to identify the landscape types with success percentage of 95% (since sensitivity is 0.95), and the success rate of detection of uncategorized Structure cases is also 95% (since specificity is 0.95). Hence the Model ESDAS is scalable and robust to predict the spatial structures of soil, forestry, valley and water resource landscapes. The prediction statistics observed from the experimental study of the ESDAS are visualized in fig1.

5 CONCLUSION

The recent contributions in contemporary literature evincing that the role of machine learning spanned to many challenges related to spatial data. The environmental data that is of the discriminative features, the relevant information retrieval strategies such as discovery of spatial and temporal patterns, exploratory analysis of spatial data, classification and decision making often relies on Machine learning algorithms. Due to the less outliers and gradual changeovers from one type to other landscape type, the information retrieval tasks like classification, pattern discovery demands deep learning of the features at machine learning tasks to achieve the accuracy at knowledge discovery. In this regard this manuscript proposed an Explorative Spatial Data Analysis Scale to perform deep learning in order to extract patterns and further using this knowledge to classify the landscape (types) with accelerated optimality. The experimental results indicating that the success classification of landscape types is 93%. This prediction rate is phenomenal since the spatial data features are highly influenced by the curse of dimensionality and variance. The future research can classify the landscape types by using evolutionary strategies with ESDAS as cost or fitness function.

REFERENCES

[1] Moglen, G. E., Eltahir, E. A., & Bras, R. L. (1998). On the sensitivity of drainage density to climate change. Water Resources Research, 34(4), 855-862.

[2] Tucker, G. E., & Bras, R. L. (1998). Hillslope processes, drainage density, and landscape morphology. Water Resources Research, 34(10), 2751-2764.

[3] Tucker, G. E., Catani, F., Rinaldo, A., & Bras, R. L. (2001). Statistical analysis of drainage density from digital terrain data. Geomorphology, 36(3), 187-202.

[4] Touya, G. (2008, June). First thoughts for the orchestration of generalisation methods on heterogeneous landscapes. In workshop on generalisation and multiple representation, Montpellier, France.

[5] Pozdnoukhov, A., &Kanevski, M. (2006). Monitoring network optimisation for spatial data classification using support vector machines. International journal of environment and pollution, 28(3-4), 465-484.

[6] Bayro-Corrochano, E., Vazquez-Santacruz, E., Moya-Sanchez, E., & Castillo-Muñis, E. (2015). Geometric Bioinspired Networks for Recognition of 2-D and 3-D Low-Level Structures and Transformations.

[7] Kanevski, M., Pozdnoukhov, A., &Timonin, V. (2009). Machine learning for spatial environmental data: theory, applications, and software. EPFL press.

[8] Friedman, J., Hastie, T., &Tibshirani, R. (2001). The elements of statistical learning (Vol. 1). Springer, Berlin: Springer series in statistics.

[9] Haykin, S., & Network, N. (2004). A comprehensive foundation. Neural Networks, 2(2004).

[10] Lee, J. A., &Verleysen, M. (2007). Nonlinear dimensionality reduction. Springer Science & Business Media.

[11] Vapnik, V. (2006). Estimation of dependences based on empirical data. Springer Science & Business Media.

[12] Foody, G. M., &Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on geoscience and remote sensing, 42(6), 1335-1343.

[13] Ham, J., Chen, Y., Crawford, M. M., &Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 43(3), 492-501.

[14] Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. International Journal of remote sensing, 23(4), 725-749.

[15] Pal, M. (2005). Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26(1), 217-222.

[16] Song, X., Duan, Z., & Jiang, X. (2012). Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image. International Journal of Remote Sensing, 33(10), 3301-3320.

[17] Waske, B., & Braun, M. (2009). Classifier ensembles for land cover mapping using multitemporal SAR imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 64(5), 450-457.

[18] Gahegan, M. (2000). On the application of inductive machine learning tools to geographical analysis. Geographical Analysis, 32(2), 113-139.

[19] Heumann, B. W. (2011). An object-based classification of mangroves using a hybrid decision tree—Support vector machine approach. Remote Sensing, 3(11), 2440-2460.