Sādhanā (2021)46:7
https://doi.org/10.1007/s12046-020-01548-2

A novel linear assorted classification method based association rule mining with spatial data

P D SHEENA SMART[1,*], K K THANAMMAL[2] and S S SUJATHA[2]

[1] (Research Scholar Registration No. 18123152282021), Department of Computer Science, S.T. Hindu College, Nagercoil, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli 627012, India
[2] Department of Computer Science, S.T. Hindu College, Nagercoil, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli 627012, India
e-mail: sheena2021@gmail.com; thanaravindran@gmail.com; sujaajai@gmail.com

MS received 30 May 2020; revised 10 November 2020; accepted 1 December 2020

Abstract. Spatial data classification and extraction is a significant problem to be resolved in data mining. The classification performance of existing techniques was not effectual to accurately mine the interesting spatial data. Furthermore, the amount of time taken for classifying the spatial data location was very higher. In order to resolve the above limitations, an Exponentiated Pareto based linear assorted classification method is introduced to reduce the incorrect classification of spatial data. This paper begins with a discussion of traditional methods of spatial data mining. The algorithm makes an association rule with spatial data objects. The technique conducts the experimental works using metrics such as classification accuracy, time complexity, space complexity and false positive rate with respect to different number of data. The proposed approach takes the forest fire dataset and El Nino dataset as input and predicts the burned area of forest fires and weather and climate conditions. The experimental results show that the proposed technique is able to increase the classification accuracy and reduce the time complexity as well as space complexity and false positive rate of spatial data mining as compared to state-of-the-art works.

Keywords. Classification; Pareto; association rule; mining; spatial data.

1. Introduction

Spatial data represents all kinds of data objects present in geographical space. Spatial data is also called geospatial data. Spatial data classification depends on the analysis of spatial objects associated with its spatial characteristics i.e. areas of the region, roads, and ponds or rivers. Besides, spatial data mining is the process of taking out implicit knowledge, spatial relations that are stored in spatial databases. The number of spatial data that are presented in a given database is sometimes very larger. Extracting valuable patterns from spatial datasets is more difficult due to the complexity of spatial data types, spatial relationships and autocorrelation. Thus, the classification and mining of spatial data are very complex. A lot of research works have been intended in conventional works to classify and mine the spatial data.

The LogitBoost ensemble-based decision tree (LEDT) method was designed in [1] for the spatial prediction of tropical forest fire vulnerability. However, the classification accuracy of this method was not enhanced. A hybrid machine learning method called RSSCART was presented in [2] for spatial prediction of landslides. But, the time complexity of spatial classification was more. A multi-scale spatial information fusion (MSIF) was introduced in [3] for mining spatial data. The MSIF minimized the computational cost. But, the parallel version of the designed method was not taken. A cognitive approach was introduced in [4] with a mixture of knowledge-based systems, integration of remote sensing image analyses, GIS analyses, shape descriptor analyses,and artificial intelligence. An integration of contamination simulations and spatial event detection model was introduced in [5]. But, the classification accuracy was not enough. A spectral-spatial method was presented in [6] with the application of a 3-D morphological profile (3D-MP) for hyperspectral data classification. However, the time complexity of this method was more. Extended Multi-Attribute Profiles (EMAPs) were introduced in [7] with the aim of taking out the spatial information. However, the computational complexity of EMAPs was not reduced.

A Map-Reduce based approach was developed in [8] to determine all co-location patterns from a spatial dataset and handling a massive amount of spatial data. But, the accuracy of spatial data mining was poor. A novel algorithm was introduced in [9] for identifying both co-location and

[*] For correspondence

Published online: 02 February 2021

segregation patterns and reducing the computational cost. However, the number of data that are incorrectly extracted was higher. An Incremental topological spatial association rule mining and clustering were performed in [10] with geographical datasets and probabilistic approaches. But, the execution time taken for spatial association rule mining was higher. However, the classification performance of existing works was not adequate due to the complexity of spatial data types, spatial relationship, and spatial correlation. To address the above mentioned existing issues in spatial data classification and mining, the Exponentiated Pareto Spatial Distributive Linear Assorted Classification (EPSDLAC) Technique is developed in this research work.

The EPSDLAC Technique improves classification accuracy by optimizing the result of working with spatial correlation information. Initially, EPSDLAC Technique takes the spatial dataset i.e. forest fires dataset as input. Then, EPSDLAC Technique proposes an Exponentiated Pareto Spatial Distribution Based Association Rules Mining (EPSD-ARM) algorithm that makes association rule with the spatial data objects to improve the classification accuracy. After constructing the spatial association rules, EPSDLAC Technique designs the Spatial Linear Assorted Model with aiming at minimizing the incorrect classification of spatial data. The Spatial Linear Assorted Model in EPSDLAC Technique uses the spatial association rules to classify each spatial data.

During the classification process, the Spatial Linear Assorted Model identifies the relationship between the spatial data and rules for efficient classification of spatial data as a forest fired region or non-forest fired region. From that, EPSDLAC Technique accurately predicts and extracts the burned region of forest fires with higher accuracy. The main contributions of the EPSDLAC Technique are as follows:

1) To enhance the performance of spatial data mining via classification as compared to state-of-the-art works, EPSDLAC Technique is proposed. The EPSDLAC Technique is introduced by integrating the Exponentiated Pareto Spatial Distribution Based Association Rules Mining and Spatial Linear Assorted Model.
2) To improve the accuracy of spatial data classification with a lower time complexity as compared to existing works, Exponentiated Pareto Spatial Distribution Based Association Rules Mining is proposed in EPSDLAC Technique. The spatial association rule is a rule that represents a certain association relationship among a set of spatial attributes for accurate classification. Quality of spatial association rules is computed by their support and confidence value. EPSDLAC technique not only considers attributes of the object but also their spatial relations.
3) To minimize the false positive rate of spatial data categorization as compared to existing works, Spatial Linear Assorted Model is proposed in EPSDLAC Technique. This model estimates the correlation between the input spatial data and generated rules to perform classification with higher accuracy.

The rest of the paper is formulated as follows. Section 2 reveals the related works. In section 3, EPSDLAC Technique is explained with the aid of the architecture diagram. In section 4, simulation settings are described and the results are discussed in section 5. Section 6 provides the conclusion of the study.

2. Related works

The Peano Count Tree (P-tree) structure was developed in [11] to obtain association rules from spatial data. P-tree offers a lossless and compressed illustration of spatial data. The P-tree structure is a space-efficient lossless data mining-ready structure for spatial data sets. But, spatial data classification accuracy was not at the required level. Random spatial-subspace clustering (RSSC) was presented in [12] to diminish the computational cost. In RSSC, a subset of data is segmented and overall solution is attained via propagation. However, the false alarm rate of spatial data extraction was more. A focal-test-based spatial decision tree (FTSDT) was designed in [13] to decrease the classification errors and training time. It utilizes new focal test approach with adaptive neighborhoods that evades oversmoothing in wedge-shaped areas. But, the classification performance of spatial data was not effective. Graph convolutional neural network (GCNN) architecture was introduced in [14] to examine graph-structured spatial vector data. This work performs classification of building pattern with design rules and extracted features for specific patterns. However, the spatial association rules were not considered in this work. A novel technique was designed in [15] to enhance the accuracy of mining spatial data related to land clearing. The spatial distribution of land cover was classified with machine learning techniques and temporal changes were considered for analysis. Supervised classification was performed with six spectral features. But, the time complexity of this technique was not solved. A cluster analysis was performed in [16] to increase the spatial data mining accuracy. However, the false positive rate remained an open issue.

Support Vector Machine-based spatial data mining was introduced in [17] for traffic risk examination. Extracting knowledge from spatial data is significant for traffic risk analysis. But, the traffic risk accuracy was poor. The P-trees algorithm was employed in [18] with the help of pruning techniques to increase the performance of the rule mining process. The association rule mining finds relationships among a set of data items. However, the classification time was more. A review of different classification and clustering algorithms designed for mining spatial data was analyzed in [19]. A recent investigation on knowledge discovery for mining interesting patterns from spatial datasets was presented in [20]. Based on the existing spatial data classification and mining methods, we encounter the challenges and provide efficient solutions. A number of algorithms have been proposed to find knowledge from spatial data. A B- Convolutional Neural Network (B-CNN) was introduced in [21] to offer high-level features for mapping flooded regions (before and after) with remote sensing data. Here, preprocessing and dimensionality reduction are performed to extract higher-level features of flooded scene matching region from satellite view. CNN is possible to train enormous data in remote sensing studies. However, classification accuracy was minimal.

An evaluation of various data mining algorithms was presented in [22] for land cover change identification. In addition, data mining techniques are able to process huge data sets and utilize the intrinsic characteristics of spatio-temporal data. The data mining technique resolves several issues in the data set through the preprocessing of data. In [23], a modified inter-layer dependency (MILD) scheme was presented. The proposed scheme includes four different processes for spatio-temporal and topologic analyses. But, the error rate was more.

3. Exponentiated pareto spatial distributive linear assorted classification

Spatial data classification is a key method in spatial data mining and spatial data analysis. Spatial data classification and spatial association rule mining attain great significance in recent years. Many research works have been designed for spatial data classification with the help of diverse data mining techniques. However, the classification accuracy of conventional works was not sufficient. Therefore, the EPSDLAC technique is developed to improve the performance of spatial data classification with higher accuracy to efficiently predict the burned area of forest fires and oceanographic and surface meteorological condition. EPSDLAC technique is proposed by combining the Exponentiated Pareto Spatial Distribution Based Association Rules Mining and Spatial Linear Assorted Model on the contrary to state-of-the-art works. In this technique, Exponentiated Pareto Distribution is very useful in describing and determining real-world phenomena. It helps the proposed EPSDLAC Technique to find recurrence relations among spatial data for efficient association rules generation. Spatial association rules are constructed to carry out the classification process with improved accuracy and minimal time complexity.

EPSDLAC technique extends the general classification methods to consider not only attributes of an object to be classified but also their spatial relations. Besides, the Spatial Linear Assorted Model is introduced in the EPSDLAC technique to classify each spatial data position with generated association rules for determining the burned area of forest fires and oceanographic and surface meteorological conditions. The architecture diagram of EPSDLAC is depicted in figure 1.

Figure 1 shows the overall processes of the EPSDLAC Technique to achieve higher classification accuracy for spatial data mining. As presented in the above figure, the EPSDLAC Technique at first gets the spatial dataset (i.e. Forest Fires Dataset [24] and El Nino Dataset [25] from UCI repository) as input. After that, EPSDLAC Technique applies the Exponentiated Pareto Distribution to formulate the spatial association rules through considering the recurrence relations between spatial data properties. Then, the EPSDLAC Technique designs Spatial Linear Assorted Model to reduce the false-positive rate of classifying the spatial data.

Thus, the EPSDLAC Technique obtains higher classification accuracy and minimal time complexity for finding a burned area of forest fires and oceanographic and surface meteorological condition as compared to state-of-the-art works. The detailed process of EPSDLAC Technique is described in below subsections.

.1. Exponentiated Pareto spatial distribution based Association Rules mining

When performing the classification process, spatial association rules are needed. A spatial association rule is a rule where one of the predicates is spatial. The quality of spatial association rules is estimated by their support and confidence. The spatial association rule mining finds certain association between collections of data items in a spatial dataset.

The Exponentiated Pareto Spatial Distribution Based Association Rules Mining (EPSD-ARM) algorithm is proposed in the EPSDLAC Technique on the contrary to existing works. The EPSD-ARM algorithm is designed to derive association rules from spatial data. To enhance the

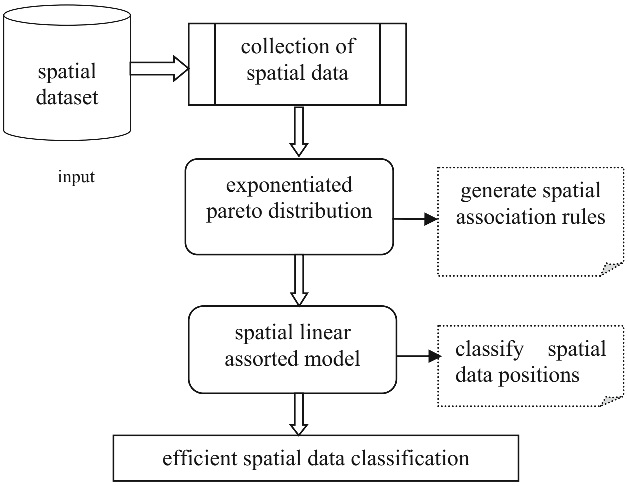
spatial dataset
input
collection of spatial data
exponentiated pareto distribution
generate spatial association rules
spatial linear assorted model
classify spatial data positions
efficient spatial data classification

Figure 1. Architecture Diagram of EPSDLAC Technique.

efficiency of the association rule mining process, Exponentiated Pareto Distribution is applied in the EPSD-ARM algorithm. By using Exponentiated Pareto Distribution, the EPSD-ARM algorithm identifies the recurrence relations among spatial data properties in the input dataset to create the association rules.

The EPSD-ARM algorithm extends the conventional associate rule mining by considering the spatial properties using Exponentiated Pareto Distribution. A spatial association rule is expressed in the form '$P \rightarrow Q[s\%, c\%]$', where '$P$' and '$Q$' are sets of spatial or non-spatial predicates. Here, '$s\%$' is the support of the rule, and '$c\%$' is the confidence of the rule. Some instances for spatial predicates are close_to, far_away, intersect, overlap, etc. which can be employed in spatial association rules.

The EPSD-ARM algorithm is employed for finding interesting relations between spatial data in an input dataset using spatial association rule to efficiently perform classification. The spatial association rules explain the spatial attribute conditions that occur frequently together in a given dataset for classifying the burned area of forest fires and oceanographic and surface meteorological conditions based on a threshold called minimum support and minimum confidence value. The support value determines the frequent item (i.e., spatial attribute) sets in an input dataset. Besides, the confidence values measure the conditional probability for finding the burned area of forest fires.

The block diagram of EPSD-ARM is depicted in figure 2 for effectively generating the spatial association rules from the given dataset.

Figure 2 shows the flow process of the EPSD-ARM algorithm that increases the efficiency of the spatial association rule mining process. Let us consider 'SD' as a spatial data set which comprises of many spatial data '$SD = \{S_1, S_2, S_3, \ldots S_N\}$'. Here, each spatial data contains a set of attributes denoted as '$\{A_1, A_2, A_3, \ldots A_M\}$'. In general, a spatial association rule is of the form '$A_1 \rightarrow A_2$' where 'A_1' and 'A_2' are disjoint spatial attributes in the given dataset. Let us assume the Exponentiated Pareto spatial distribution '$EPSD(\theta, \lambda)$' with probability density function which is formulated as,

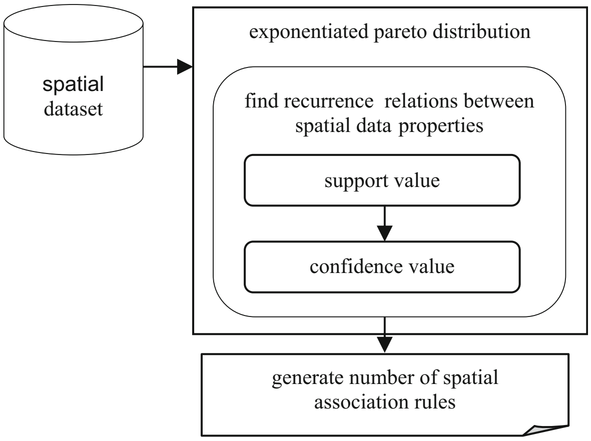
spatial
dataset
exponentiated pareto distribution
find recurrence relations between
spatial data properties
support value
confidence value
generate number of spatial
association rules

Figure 2. EPSD-ARM Process for Spatial Association Rules Generation.

$$f(S_i) = \theta\lambda\left[1 - (1 + S_i)^{-\lambda}\right]^{\theta-1}(1 + S_i)^{-(\lambda+1)}, S_i > 0, \lambda > 0, \theta > 0 \tag{1}$$

From (1), 'S_i' indicates spatial data. In the same way, Exponentiated Pareto spatial distribution '$EPSD(\theta, \lambda)$' with cumulative distribution function is mathematically obtained as,

$$F(S_i) = \left[1 - (1 + S_i)^{-\lambda}\right]^{\theta}, \quad S_i > 0, \lambda > 0, \theta > 0 \tag{2}$$

In the above equations (1) and (2), 'θ' and 'λ' are two shape parameters. Here, 'S_i' represents the 'i^{th}' spatial data and its attributes in the given dataset. Let us consider '$S_1, S_2, S_3, \ldots S_N$' be a sequence of random spatial data with probability density function '$f(S_i)$' and cumulative distribution function '$F(S_i)$', then recurrence relations among the spatial data attribute in a given dataset is mathematically determined as,

$$\mu_m^{(S_i)} = \frac{S_i}{\theta\lambda}\sum_{i=2}^{\lambda+1}\binom{\lambda+1}{i}\mu_{m+1}^{(S_i+i-1)} + \left(\frac{S_i}{\theta} + 1\right)\mu_{m+1}^{(S_i)}, \tag{3}$$

where $S_i = 1, 2, \ldots.$

From equation (3), recurrence relations for each spatial data 'S_i' in an input dataset is determined. From the above equation, we find the relation for entire spatial data. Based on measured recurrence relations, then the support value is mathematically obtained using below,

$$Sup(A_1 \rightarrow A_2) = \left(\frac{Support\ count\ of\ A_1 \cup A_2}{Total\ number\ of\ attributes\ in\ spatial\ dataset}\right) \tag{4}$$

From equation (4), the support value is determined using the number of spatial data that contain '$A_1 \cup A_2$' to the total number of spatial data in a given dataset. Here, we find the relation for two different spatial data. Conversely, the confidence value is evaluated in such a way that the transactions which include 'A_1' also contain 'A_2' using the below mathematical expression,

$$Conf(A_1 \rightarrow A_2) = \left(\frac{Support\ count\ of\ A_1 \cup A_2}{Support\ count\ of\ A_1}\right) \tag{5}$$

From eqs. (4) and (5), the support and confidence value is measured for each spatial data attributes in a given dataset. EPSD-ARM algorithm then defines a minimum threshold for support and confidence value. Finally, the attributes which contain minimum support and confidence are selected to make the spatial association rules. The algorithmic steps of EPSD-ARM are explained in Algorithm 1.

Algorithm 1 depicts the step by step process of the EPSD-ARM to improve the performance of spatial association rule mining. With the help of the above algorithmic process, the EPSD-ARM algorithm efficiently constructs the spatial association rules with a minimal amount of time by considering the recurrence relations among spatial attributes in the input dataset.

3.1 *Spatial linear assorted model*

The Spatial Linear Assorted Model is designed to increase the classification performance of spatial data by using constructed spatial association rules. In EPSDLAC Technique, the Spatial Linear Assorted Model determines the relationship between association rules and spatial data for efficient classification. In addition to that, the Spatial Linear Assorted Model considers the covariate measurement error and spatial correlation to minimize the incorrect classification of spatial data on the contrary to conventional works. The process involved in the Spatial Linear Assorted Model is depicted in figure 3.

Figure 3 shows the flow process of the Spatial Linear Assorted Model. This model at first takes a number of spatial data and constructs association rules as input. Followed by this, the Spatial Linear Assorted Model measures the relationship between the rules and spatial data. Based on the determined relationships, then the Spatial Linear Assorted Model classifies the spatial data with higher accuracy and a lower amount of time consumption. Let us consider a spatial dataset comprised of numerous spatial data represented as ‘$SD = S_1, S_2, S_{3,\ldots,}, S_N$’ and generated spatial association rules denoted as ‘$R_1, R_2, R_3, \ldots R_N$’. Consider that the spatial data are obtained from ‘n’ geographical areas. The spatial linear mixed model of ‘Y’ for the given spatial data ‘S_i’ and spatial association rules ‘R_i’ is mathematically determined as,

$$Y_i = \beta_0 + S_i\beta_i + R_i\beta_i + b_i + \varepsilon_i \tag{6}$$

From the above equation (6), random effects ‘β_i’ model the spatial correlation. Here, the random effect vector ‘$(\alpha_1, \ldots, \alpha_n)$’ is ‘$n\{0, \tau(\theta)\}$’ and ‘$\theta$’ is a vector of variance components, the residuals ‘ε_i’ is ‘$n(0, \sigma_\varepsilon^2)$’, and ‘ b_i’ and ‘ ε_i’are independent to each other and are independent of the covariates ‘S_i’ and ‘R_i’. The naive estimators of $(\beta_0, \beta_i, \theta)$ are obtained by simply replacing ‘S_i’ with the error-prone observation and fitting. In EPSDLAC Technique, the covariance matrix ‘$\tau(\theta)$’ models the connection between the spatial data ‘S_i’ and association rules ‘R_i’ to accurately classify the data. The algorithmic steps of the Spatial Linear Assorted Model are explained in Algorithm 2.

Algorithm 2. Spatial Linear Assorted Model.

```
Input:   Spatial Dataset ‘$SD$’, Collections of Spatial
         Data ‘$SD = \{S_1, S_2, S_3, \ldots S_N\}$’, set of Spatial
         Attributes denoted as ‘$\{A_1, A_2, A_3, \ldots A_M\}$’.
         Spatial Association Rules ‘$R_1, R_2, R_3, \ldots R_N$’
Output:  Higher spatial data classification with
         minimum time
Step 1: Begin
Step 2:   For Each Spatial Data ‘$S_i$’
Step 3:     Apply spatial association rules ‘$R_i$’
Step 4:     Measure the correlation between ‘$S_i$’ and
            ‘$R_i$’ using (6)
Step 5:     Classify ‘$S_i$’
Step 6:     Extract the classified spatial data
Step 7:   End for
Step 8: End
```

Algorithm 2 presents the step by step process of the Spatial Linear Assorted Model to reduce the error rate of spatial data classification. As explained in the above algorithmic steps, Spatial Linear Assorted Model initially gets a number of spatial data and association rule as input. Consequently, Spatial Linear Assorted Model finds out the relationship between the spatial data and rules to effectively classify the data with improved accuracy and minimal time. As a result, EPSDLAC Technique increases the accuracy and minimizes the time of spatial data classification and mining when compared to state-of-the-art works.

4. Experimental settings

In order to evaluate the performance of the proposed work, the EPSDLAC Technique is implemented in Java Language using Forest Fires Dataset [24] and El Nino Dataset [25] obtained from the UCI machine learning repository. This Forest Fires Dataset comprises 517 spatial data and 13 spatial attributes to find the burned area of forest fires. El Nino dataset comprises oceanographic and surface meteorological readings from sequence of buoys placed in the equatorial Pacific. The data assist to predict the El Nino/Southern Oscillation (ENSO) cycles. The dataset characteristics are spatio-temporal and attribute characteristics are both real and integer. Number of instances are 178080 and number of attributes are 12. The spatial attributes information of Forest Fires Dataset is depicted in table 1.

The spatial attributes information of using El Nino dataset is shown in table 2.

The EPSDLAC Technique considers the 50-500 numbers of spatial data from the Forest Fires dataset and 10-100

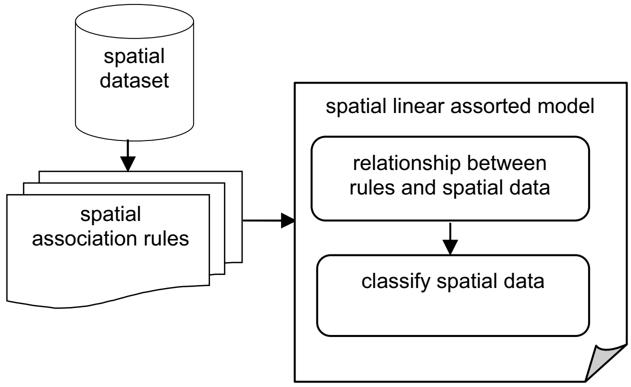
spatial dataset
spatial association rules
spatial linear assorted model
relationship between rules and spatial data
classify spatial data

Figure 3. Block Diagram of Spatial Linear Assorted Model for Spatial Data Classification.

numbers of spatial data from the El Nino dataset to conduct the experimental evaluation.

The performance of the EPSDLAC Technique is determined in terms of classification accuracy, time complexity, space complexity and false positive rate with respect to various numbers of spatial data.

The experimental result of the proposed EPSDLAC Technique is compared against two conventional methods namely LogitBoost ensemble-based decision tree (LEDT) [1] and a hybrid machine learning approach of Random Subspace (RSS) and Classification And Regression Trees (CART) called RSSCART [2].

4.1 *Case study*

Experimental evaluation of EPSDLAC Technique is implemented in Java Language with two different datasets namely Forest Fires Dataset [24] and El Nino dataset [25]. Initially, we consider 50 numbers of spatial data from Forest Fires Dataset. Then, Exponentiated Pareto Spatial Distribution Based Association Rules mining is applied which denotes a certain association relationship among set of spatial attributes for accurate classification. With the 50 number of spatial data, 84% of classification accuracy is achieved in the proposed technique. Whereas, the classification accuracy using El Nino dataset is 92% when considering 100 spatial data. With this, the proposed EPSDLAC Technique will be able to exactly predict the objects based on the spatial data.

EPSDLAC Technique finds the burned area of forest fire and oceanographic and surface meteorological condition based on threshold called minimum support and minimum confidence value. It also utilizes longitude and latitude data to exactly identify the objects.

5. Results and discussion

In this section, a comparative result analysis of the proposed EPSDLAC Technique is presented. The effectiveness of the EPSDLAC Technique is compared with existing LEDT [1] and RSSCART [2] using parameters such as classification accuracy, time complexity, space complexity and false-positive rate with the help of below tables and graphic representation.

5.1 *Classification accuracy*

The Classification Accuracy '*CA*' is determined as the ratio of a number of spatial data correctly classified to the total number of spatial data taken as input. The classification accuracy is evaluated in terms of percentage (%) and formulated as below,

$$CA = \frac{X_C}{N} * 100 \tag{7}$$

From equation (7), the accuracy of spatial data classification is measured. Here, '*N*' refers to the total number of

Table 1. Spatial Attributes Information using forest fires dataset information.

Attribute	Description
X	x-axis spatial coordinate within the Montesinho park map: 1 to 9
Y	y-axis spatial coordinate within the Montesinho park map: 2 to 9
Month	the month of the year: 'jan' to 'dec'
Day	day of the week: 'mon' to 'sun'
FFMC	FFMC index from the FWI system: 18.7 to 96.20
DMC	DMC index from the FWI system: 1.1 to 291.3
DC	DC index from the FWI system: 7.9 to 860.6
ISI	ISI index from the FWI system: 0.0 to 56.10
Temp	the temperature in Celsius degrees: 2.2 to 33.30
RH	relative humidity in %: 15.0 to 100
Wind	wind speed in km/h: 0.40 to 9.40
Rain	outside rain in mm/m2 : 0.0 to 6.4
Area	the burned area of the forest (in ha): 0.00 to 1090.84

Table 2. Spatial Attributes Information using El Nino dataset information.

Attribute	Description
date	Date of the readings taken
Latitude	Values stayed within a degree from the approximate location
Longitude	Values were sometimes as far as five degrees off of the approximate location.
Zonal winds (west<0, east>0)	zonal winds are fluctuated between -10 m/s and 10 m/s
Meridional winds (south<0, north>0)	Meridional winds fluctuated between -10 m/s and 10 m/s
Relative humidity	Relative humidity values in the tropical Pacific were typically between 70% and 90%
Air temperature	Air temperature fluctuated between 20 and 30 degrees Celsius
Sea surface temperature	Sea surface temperature fluctuated between 20 and 30 degrees Celsius
Subsurface temperatures	Subsurface temperatures down to a depth of 500 meters

spatial data employed for conducting experimental work, whereas 'X_C' points out the number of spatial data that are accurately classified.

Tables 3 and 4 illustrate the classification accuracy using two different datasets. From the above tables 3 and 4, the classification accuracy of EPSDLAC is higher. The graphical representation of classification accuracy is given below.

Figure 4 depicts the impact of classification accuracy using Forest Fires Dataset versus a different number of spatial data using three methods namely LEDT [1], RSSCART [2] and the Proposed Technique. The results of classification accuracy using El Nino Dataset are portrayed in figure 5.

As illustrated in the above figures, EPSDLAC Technique gets enhanced classification accuracy for spatial data to discover the burned area of forest fires and oceanographic and surface meteorological conditions when compared to existing LEDT [1] and RSSCART [2]. This is owing to the application of the Exponentiated Pareto Spatial Distribution Based Association Rules Mining and Spatial Linear Assorted Model in the proposed EPSDLAC Technique on contrary to conventional works.

During the rule generation process, EPSDLAC Technique provides the best spatial association rules for improving classification performance with the aid of the EPSD-ARM algorithm. During the classification process, the EPSDLAC Technique effectively classifies the spatial data into a corresponding class such as forest fired region or non-forest fired region with the help of the Spatial Linear Assorted Model. This supports the EPSDLAC Technique to improve the ratio of a number of spatial data properly classified as compared to state-of-the-art works.

With the Forest Fires Dataset, the EPSDLAC Technique enhances the classification accuracy by 20% and 14% when compared to LEDT [1] and RSSCART [2] respectively. By utilizing the El Nino dataset, the classification accuracy of EPSDLAC Technique is enhanced by 18% and 12% as compared to LEDT [1] and RSSCART [2].

Table 3. Tabulation for Classification Accuracy (%) using Forest Fires Dataset.

	Classification Accuracy (%)		
Number of spatial data (N)	LEDT	RSSCART	EPSDLAC
50	66	70	84
100	68	77	91
150	75	82	93
200	80	80	91
250	82	86	95
300	82	85	93
350	81	84	95
400	79	81	93
450	77	81	90
500	77	80	89

Table 4. Tabulation for Classification Accuracy (%) using El Nino Dataset.

	Classification Accuracy (%)		
Number of spatial data (N)	LEDT	RSSCART	EPSDLAC
10	50	54	60
20	55	58	65
30	60	63	70
40	62	65	72
50	65	68	75
60	68	72	80
70	70	75	82
80	72	77	88
90	75	79	90
100	78	79	92

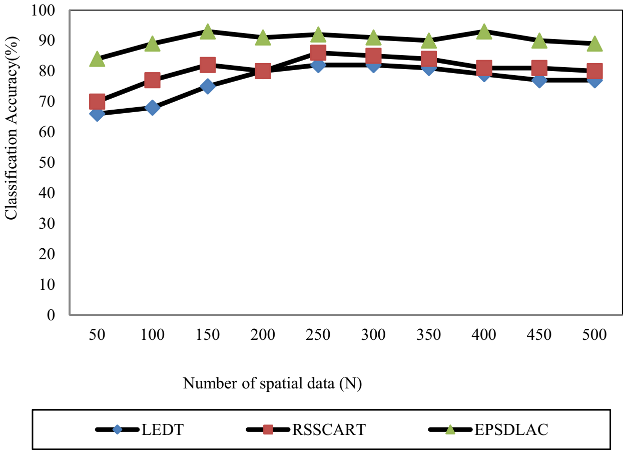

Figure 4. Experimental Result Analysis of Classification Accuracy Vs Number of Spatial Data using Forest Fires Dataset.

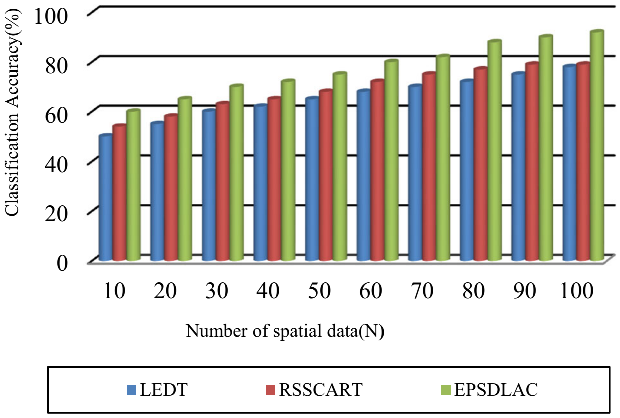

Figure 5. Experimental Result Analysis of Classification Accuracy Vs Number of Spatial Data using El Nino Dataset.

5.2 *Time complexity*

The Time Complexity '(TC)' measures the amount of time needed to classify the spatial data.

The time complexity is estimated in terms of milliseconds (ms) and obtained using below,

$$TC = N * t(\mathrm{C_{SD}}) \tag{8}$$

From equation (8), the time complexity of spatial data classification is determined. Here, '(N)' denotes the number of spatial data, and '$t(C_{SD})$' represents the time utilized for classifying a single spatial data. For determining the time complexity involved during the spatial data classification process, the EPSDLAC Technique is implemented in Java language with different numbers of spatial data in the range of 50-500 from the forest fires dataset.

The experimental result of time complexity using the EPSDLAC Technique is compared with two state-of-the-art methods LEDT [1] and RSSCART [2]. When taking 400 spatial data to conduct the simulation process, EPSDLAC Technique obtains 60 ms time complexity whereas existing LEDT [1] and RSSCART [2] gets 76 ms and 68 ms respectively. The tabulation result analysis of time complexity using Forest Fires Dataset is demonstrated in table 5.

Table 6 portrays the result of time complexity using El Nino dataset. From tables 5 and 6, it is obvious that EPSDLAC Technique obtains minimal time complexity for spatial data classification and thereby determines the burned area of forest fires and oceanographic and surface meteorological condition as compared to existing LEDT [1] and RSSCART [2]. This is due to the application of the Exponentiated Pareto Spatial Distribution Based Association Rules Mining and Spatial Linear Assorted Model in the proposed EPSDLAC Technique on contrary to existing works. With the support of the EPSD-ARM algorithm process, EPSDLAC Technique makes the best spatial association rules with a minimal amount of time. Besides that, the EPSDLAC Technique accurately categories the spatial data into a corresponding class with a lower amount of time utilization. This helps the EPSDLAC Technique to decrease the amount of time required to classify the spatial data when compared to state-of-the-art works. Therefore, EPSDLAC Technique reduces the time complexity of spatial data classification using Forest Fires Dataset by 35% and 21% when compared to LEDT [1] and RSSCART [2] respectively. By applying El Nino dataset, EPSDLAC Technique minimizes the time complexity by 35% and 30% when compared to LEDT [1] and RSSCART [2] respectively.

Figures 6 and 7 depicts the Time complexity results of three different methods using two datasets. From the above figure, EPSDLAC Technique attains minimal time complexity for spatial data classification as compared to existing LEDT [1] and RSSCART [2].

Table 5. Tabulation for Time Complexity (ms) using Forest Fires Dataset.

	Time Complexity (ms)		
Number of spatial data (N)	LEDT	RSSCART	EPSDLAC
50	19	16	12
100	30	23	19
150	44	33	26
200	54	40	30
250	65	50	35
300	63	57	36
350	70	56	46
400	76	68	60
450	81	72	66
500	100	85	75

Table 6. Tabulation for Time Complexity (ms) using El Nino Dataset.

	Time Complexity (ms)		
Number of spatial data (N)	LEDT	RSSCART	EPSDLAC
10	25	23	19
20	28	22	17
30	26	24	15
40	22	20	13
50	21	19	14
60	27	24	16
70	30	29	20
80	35	35	23
90	38	40	25
100	48	45	30

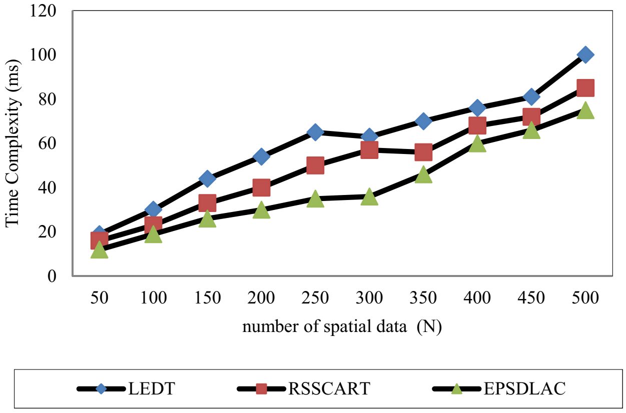

Figure 6. Experimental Result Analysis of Time Complexity Vs Number of Spatial Data using Forest Fires Dataset.

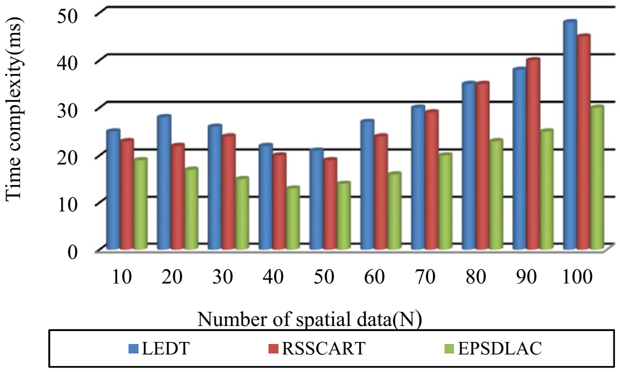

Figure 7. Experimental Result Analysis of Time Complexity Vs Number of Spatial Data using El Nino Dataset.

5.3 *Space complexity*

Space complexity is measured as the amount of memory space required for storing the spatial data. The Space complexity is formulated as follows,

$$SC = Numberofdata * space(storingonedata) \qquad (9)$$

From equation (9), space complexity is obtained. Here, *SC* indicates Space complexity. It is measured in terms of megabytes (MB).

Table 7 illustrates the results of space complexity using Forest Fires Dataset with a number of spatial data in the range of 50 to 500. Table 8 portrays the space complexity results using El Nino dataset with number of spatial data ranges from 10 to 100. From tables 7 and 8, it is evident that the space complexity using EPSDLAC is very minimal. When considering the 500 spatial data from Forest Fires Dataset, the space complexity of EPSDLAC is 20 MB, whereas, the space complexity of LEDT [1] and RSSCART [2] is 25 MB and 27 MB. When considering the 50 number of spatial data from El Nino dataset, the space complexity of EPSDLAC is 20 MB, whereas, the space complexity of LEDT [1] and RSSCART [2] is 31 MB and 29 MB. This is due to the application of Exponentiated Pareto Spatial Distribution Based Association Rules Mining and Spatial Linear Assorted Model in EPSDLAC. The space complexity of EPSDLAC using Forest Fires Dataset is reduced by 27% and 36% as compared to LEDT [1] and RSSCART [2] respectively. By applying El Nino Dataset, the space complexity of EPSDLAC is minimized by 28% and 25% when compared to LEDT [1] and RSSCART [2] respectively.

Figures 8 and 9 depict the impact of space complexity with respect to a diverse number of spatial data using two dataset. From figures 8 and 9, it is obvious that EPSDLAC Technique obtains minimal space complexity as compared to existing LEDT [1] and RSSCART [2].

5.4 *False positive rate*

False Positive Rate '(*FPR*)' evaluates the ratio of the number of spatial data wrongly classified as positive and

Table 7. Tabulation for Space complexity (MB) using Forest Fires Dataset.

	Space Complexity (MB)		
Number of spatial data (N)	LEDT	RSSCART	EPSDLAC
50	18	21	13
100	20	23	11
150	22	24	14
200	21	25	12
250	19	23	15
300	17	21	14
350	20	25	17
400	18	20	12
450	22	23	19
500	25	27	20

Table 8. Tabulation for Space complexity (MB) using El Nino Dataset.

	Space Complexity (MB)		
Number of spatial data (N)	LEDT	RSSCART	EPSDLAC
10	30	28	21
20	33	30	22
30	35	32	23
40	32	31	25
50	31	29	20
60	29	33	19
70	34	27	27
80	36	28	24
90	32	34	22
100	28	36	25

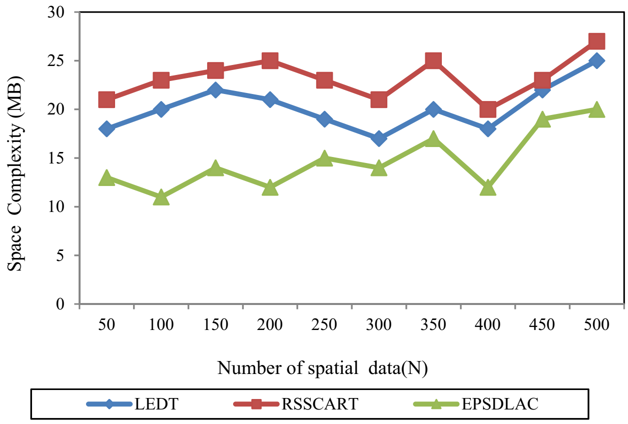

Figure 8. Experimental Result Analysis of space complexity Vs Number of Spatial Data using Forest Fires Dataset.

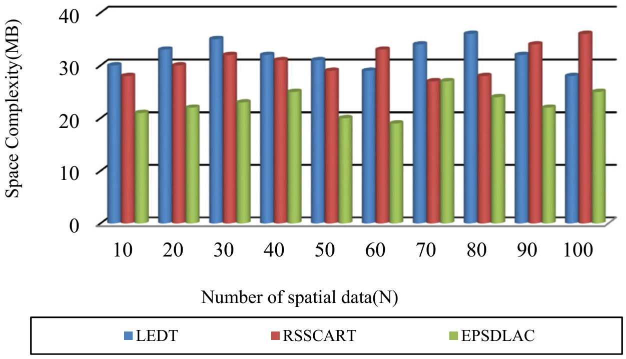

Figure 9. Experimental Result Analysis of space complexity Vs Number of Spatial Data using El Nino Dataset.

the total number of actual negative events. The false positive rate is measured in terms of percentage (%) and formulated using below,

$$FPR = \frac{FP}{FP + TN} * 100 \tag{10}$$

From equation (10), the false positive rate of spatial data classification is obtained. Here, '*FP*' denotes the False positive and '*TN*' denotes the true negative.

Tables 9 and 10 illustrate the experimental result of False Positive Rate using the EPSDLAC Technique which is compared with two state-of-the-art methods LEDT [1] and RSSCART [2] using two datasets.

Figures 10 and 11 portray the impact of false positive rate based on a varied number of spatial data using three methods namely LEDT [1] and RSSCART [2] and proposed Technique with two different datasets. As shown in the above figures, EPSDLAC Technique attains minimal false positive rate to accurately classify the spatial data and thereby find out the burned area of forest fires and oceanographic and surface meteorological condition when compared to existing LEDT [1] and RSSCART [2]. This is because of the application of the Spatial Linear Assorted Model in the proposed EPSDLAC Technique on the contrary to state-of-the-art methods where it finds the correlation between the rules and spatial data. Based on the result of this correlation, the Spatial Linear Assorted Model categorizes the spatial data as a forest fired region or non-forest fired region with enhanced accuracy. This assists in the EPSDLAC Technique to minimize the ratio of the number of spatial data incorrectly classified when compared to state-of-the-art works. Therefore, EPSDLAC Technique decreases the false positive rate of spatial data classification using Forest Fires Dataset by 57% and 49% when compared to LEDT [1] and RSSCART [2], respectively. By applying the El Nino dataset, EPSDLAC Technique decreases the false positive rate by 15% and 13% as compared to LEDT [1] and RSSCART [2].

Table 9. Tabulation for false positive rate (%) using Forest Fires Dataset.

	False Positive Rate (%)		
Number of spatial data (N)	LEDT	RSSCART	EPSDLAC
50	34	30	16
100	32	23	11
150	25	18	7
200	21	20	9
250	18	14	8
300	18	15	9
350	19	16	10
400	22	19	8
450	23	19	10
500	23	20	11

Table 10. Tabulation for false positive rate (%) using El Nino Dataset.

	False Positive Rate (%)		
Number of spatial data (N)	LEDT	RSSCART	EPSDLAC
10	30	32	25
20	35	35	30
30	38	36	33
40	40	38	35
50	43	43	39
60	47	45	40
70	50	47	42
80	54	50	48
90	57	53	43
100	60	59	45

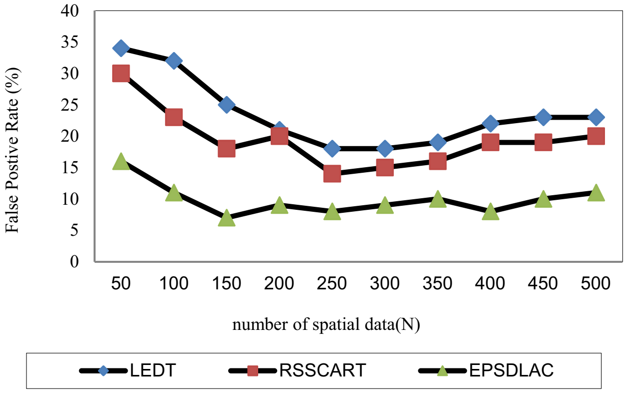

Figure 10. Experimental Result Analysis of False Positive Rate Vs Number of Spatial Data using Forest Fires Dataset.

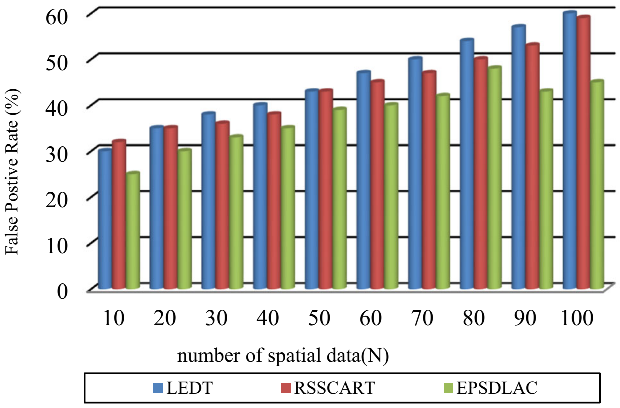

Figure 11. Experimental Result Analysis of False Positive Rate Vs Number of Spatial Data using El Nino Dataset.

6. Conclusion

An effective EPSDLAC Technique is designed to increase the spatial data mining performance through classification. The goal of the EPSDLAC Technique is attained with the application of the Exponentiated Pareto Spatial Distribution Based Association Rules Mining and Spatial Linear Assorted Model. The designed EPSDLAC Technique gets the best spatial association rules for spatial data classification as compared to conventional works by the application of the EPSD-ARM algorithm.

Moreover, EPSDLAC Technique takes a minimal amount of time to classify and mine spatial data as compared to state-of-the-art works with the help of formulated spatial association rules. In addition to that, EPSDLAC Technique decreased the ratio of the number of spatial data wrongly classified for efficient spatial data mining as compared to state-of-the-art methods with the application of the Spatial Linear Assorted Model. Thus, EPSDLAC Technique enhances the performance of spatial data mining as compared to existing works.

The efficiency of the EPSDLAC Technique is evaluated in terms of classification accuracy, time complexity, space complexity and false positive rate and compared with two existing methods. The simulations results demonstrate that EPSDLAC Technique presented a better performance with an enhancement of classification accuracy and reduction of time complexity as well as space complexity and false positive rate to extract the spatial information when compared to the state-of-the-art works.

References

[1] Tehrany M S, Jones S, Shabani F, Martínez-Álvarez F and Bui D T 2018 A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. *Theor. Appl. Climatol,* 137: 637-653, 1-17

[2] Pham B T, Prakash I and Bui D T 2017 Spatial Prediction of Landslides Using Hybrid Machine Learning Approach Based on Random Subspace and Classification and Regression Trees. *Geomorphology* 303: 256-270

[3] Li H, Song Y and Chen C L P 2017 Hyperspectral Image Classification based on Multiscale Spatial Information Fusion. *IEEE Trans. Geosci Remote Sens.* 55: 5302-5312

[4] Ojaghi S, Ahmadi F F and Ebadi H 2016 A new method for semi-automatic classification of remotely sensed images developed based on the cognitive approaches for producing spatial data required in geomatics applications. *Arab. J. Geosci.,* 9: 1-12

[5] Oliker N, Ohar Z and Ostfeld A 2016 Spatial event classification using simulated water quality data. *Environ. Model. Softw.* 77: 71-80

[6] Hou B, Huang T, Jiao L 2015 Spectral-Spatial Classification of Hyperspectral Data Using 3-D Morphological Profile. *IEEE Geosci. Remote. Sens. Lett.* 12: 1-5

[7] Ghamisi P, Benediktsson J A, Cavallaro G and Plaza A 2014 Automatic Framework for Spectral-Spatial Classification Based on Supervised Feature Extraction and Morphological Attribute Profiles. *IEEE J. Sel. Top Appl. Earth Obs. Remote* Sens. 7: 2147 – 2160

[8] Maiti S and Subramanyam R B V 2018 Mining co-location patterns from distributed spatial data, *J. King Saud Univ., Comp. & Info. Sci.* pp. 1-10

[9] Barua S and Sander J 2014 Mining Statistically Significant Co-location and Segregation Patterns. *IEEE Trans. Knowl. Data Eng*. 26: 1185-1199

[10] Jayababu Y, Varma G P S and Govardhan A 2017 Incremental topological spatial association rule mining and clustering from geographical datasets using probabilistic approach. *J. King Saud Univ., Comp. & Info. Sci.* 30: 510-523

[11] Ding Q, Ding Q and Perrizo W 2008 PARM—An Efficient Algorithm to Mine Association Rules From Spatial Data. *IEEE Trans. Syst*., *Man*, *Cybern. B. Cyber,* 38: 1513-1524

[12] Guo Y, Gao J, Li F 2015, Random spatial-subspace clustering. *Knowl. Based Syst.* 74: 106-118

[13] Jiang Z, Shekhar S, Zhou X, Knight J and Corcoran J 2015 Focal-Test-Based Spatial Decision Tree Learning. *IEEE Trans. Knowl. Data Eng*. 27: 1547 – 1559

[14] Yan X, Ai T, Yang M and Yin H 2019 A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS J. Photogramm. Remote Sens.* 150: 259-273

[15] Vasuki Y, Yu L, Holden E, Kovesi P, Wedge D and Grigg A H 2018 The spatial-temporal patterns of land cover changes due to mining activities in the Darling Range, Western Australia: A Visual Analytics Approach. *Ore Geol. Rev.* 1-41

[16] Kumar C N S, Ramulu V S, Reddy K S, Kotha S and Kumar C M 2012 Spatial Data Mining using Cluster Analysis. *International Journal of Computer Science & Information Technology*, 4:71-77

[17] Kumar R, Chundawat D S and Singh P K 2014 SVM Based Spatial Data Mining for Traffic Risk Analysis. *International Journal of Engineering Research and General Science*, 2: 716-718

[18] Sananse D A and Tuteja R R 2015 Association Rules Mining Technique Based on Spatial Data Classification. *International Journal of Computer Science Engineering and Technology*, 5: 131-136

[19] Lakumarapu S and Agarwal D R 2016, Classification of Spatial Data Mining Algorithm by Clustering Technique. *International Journal of Electronics, Electrical and Computational System*, 5: 59-64

[20] Hemalatha D M and Saranya N N 2011 A Recent Survey on Knowledge Discovery in Spatial Data Mining. *International Journal of Computer Science Issues*, 8: 473-479

[21] Banupriya R and Kannan A R 2020 Satellite image based flood classification in urban areas using B-convolutional networks. *Sādhanā*, 45: 1-5

[22] Panigrahi S, Verma K and Tripathi P 2017 Data mining algorithms for land cover change detection: a review. *Sādhanā*, 42: 2081–2097

[23] Jadhav P and Kshirsagar S 2016 Efficient rate control scheme using modified inter-layer dependency for spatial scalability. *Sādhanā*, 41: 1415–1424

[24] Cortez P and Morais A 2007 *Forest Fires Dataset* [dataset]. UCI Machine Learning Repository https://archive.ics.uci.edu/ml/datasets/forest+fires

[25] Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, US department of Commerce 1999 *El Nino dataset* [dataset]. UCI Machine Learning Repository, https://archive.ics.uci.edu/ml/datasets/El+Nino