Journal of AgriSearch, 6(3):134-138
An Open Access International Peer Reviewed Quarterly
ISSN : 2348-8808 (Print), 2348-8867 (Online)
https://doi.org/10.21921/jas.v6i03.16215

# Spatial Analysis of Water Resources Data in Selected Districts of Bihar

MANIBHUSHAN* AND AKRAM AHMED

**ARTICLE INFO**

| | | |
|---|---|---|
| Received on | : | 06-70-2019 |
| Accepted on | : | 20-08-2019 |
| Published online | : | 01-09-2019 |

## ABSTRACT

The main aim of this study is to apply geographic information system (GIS) and data mining techniques to get the attribute data in a spatial and tabular form related to district wise availability of standing water bodies in their area and number of Bihar state. An analysis has been done on available spatial data and maps to get non-spatial/ tabular data, which are in a more easily understandable form. Data extracted district-wise related to area and number of standing water bodies according to their size of Bihar state. Study shows that the number and area of standing water bodies in Madhubani, East Champaran and Patna districts are 2185, 1753, 350 and 2355.42, 6752.36 and 8429.68 ha respectively. In this way, number and area of standing water bodies of other districts of Bihar are also extracted from geodatabases and digitized maps. This type of information is more useful than the spatial data because a common person is able to understand these tabular data and they can use this data for their own purposes. These data can be utilized by scientific personnel as well as farmers and that will be used in agriculture for better utilization of water resources to enhance agricultural productivity and income of farmers of Bihar state.

*KEYWORDS*

Spatial data, tabular data, GIS, data mining, Bihar

## INTRODUCTION

Data mining is the process of extracting important and useful information from a large set of data (Abello *et al*., 2002).The goal of the data mining process is to extract information from an existing data set and transform it into a human-understandable form for its better use. A spatial database is a database which has been specially optimized to store data pertaining to objects in the real world. In other words, spatial data is the data which represents objects in geometric space. The objects are stored in a database in the form of lines, points and polygons. A Relational Database Management System (RDBMS) with additional features can support spatial databases which are extensively used in environmental studies, Global Positioning System (GPS), and Geographic Information System (GIS). Spatial Data Mining (SDM) is a process of discovering trends or patterns from large spatial databases that hold geographical data. Objects in space such as lakes, ponds, rivers, forests, buildings, cities, etc., are stored in a spatial database. Spatial databases are so complex and make the SDM more difficult when compared with traditional databases. GIS, Microsoft SQL, Oracle, etc., are used for building spatial and attribute/ tabular/ non-spatial databases. Generally, a GIS is used to store, retrieve and manipulate spatial data. Spatial data mining has many applications in extracting data in agriculture. Spatial data mining can be used in GIS pertaining to railways. Various techniques werestudied that can be used in making a Railway Geographic Information System (RGIS) which can be used for spatial data presentation and statistical analysis besides helping in making well-informed decisions (Xu *et al*., 2003). Some of the techniques useful for RGIS include spatial analysis, induction, classification and clustering, trend or spatial characteristic analysis, pattern recognition and digital map image analysis. Other approaches that can also be applied to spatial data mining are visualization approaches, rough set and fuzzy set approaches, genetic algorithms, and artificial neural networks. Clustering area geographical entities were studied by using clustering algorithms (Guang-xue *et al*., 2010). The algorithms work on the concept of geometric space similarity. Artificial neural network was applied to forest resource management in the aspect of insect pest prediction using various layers of geographical data for training models (Peng and Wen 1999). Moreover, the techniques for mining spatial data are to reveal exceptional phenomena implicitly existing inside a given set of geographical data, which are informative in making decisions. Crop yield prediction has important utility towards precision agriculture. Towards this, many researchers proposed techniques. Many techniques came into existence using remote sensing data that are linked to crop yield prediction directly or indirectly (Prasad *et al*., 2006). The techniques include Temperature Condition Index (TCI), Vegetation Condition Index (VCI) and Normalized Vegetation Index (NDVI).A data warehouse as a storehouse/ big database is a repository of data collected from multiple data sources and is intended to be used as a whole under the same unified schema. A data warehouse gives the option to analyze data from different sources under the same platform.

Spatial databases are databases that store geographical information like digitized maps, and global or regional positioning. Such spatial databases present a challenge to extract data using data mining techniques. In this research, a number of digitized maps related to water resources were collected from different sources. After that, data mining techniques were used to extract non-spatial data in tabular form, which are more understandable to a common person and farmers also. Some digitized maps/ spatial data have also been created as required for analysis purpose. Relational properties of database using SQL, ANN, GIS, etc. have been

Division of Land & Water Management, ICAR Research Complex for Eastern Region, Patna, Bihar, India
*Corresponding author email: mani_patna2000@yahoo.com

used for data mining and analysis of data. In the present research work and paper, only standing water bodies data in their area and number of all districts of Bihar have extracted from digitized maps using some data mining techniques. Data mining is the process of discovering previously unknown and potentially interesting patterns in large datasets (Piatetsky-Shapiro and Frawley 1991). The 'mined' information is typically represented as a model of the semantic structure of the dataset, where the model may be used on new data for prediction or classification (Cunningham and Holmes 1999). Data mining, also termed as knowledge discovery, is the process of analyzing data from different perspective and summarizing it into valuable or non-trivial information. This information can be used for variety of purposes-research. The data can be analyzed from many different dimensions, categorized & summarized the relationships identified. Technically, data mining simply is the process of finding correlation or patterns among dozens of field in large RDBMS (Ansari and Ansari 2010). Data mining (DM) is a high-level application technique used to present and analyze data for decision-makers.

There is an enormous wealth of information embedded in huge databases belonging to enterprises and this has spurred tremendous interest in areas of knowledge discovery and data mining. Agricultural and biological research studies have used various techniques of data analysis including, natural trees, statistical machine learning and analysis methods (Cunningham and Holmes 1999). The aim of this research is to extract data in non spatial/ tabular form of all districts of Bihar using different data mining techniques and their applications to agricultural-related areas. Spatial data/ maps have been used for this purpose. Some required digitized maps of available standing water bodies of some districts of Bihar have also been created using GIS. Relation properties of the database, SQL, OLAP ( On-Line Analytical Processing), Artificial Neural Network (ANN), etc. have been used for data mining of district wise standing water bodies in their number and area of Bihar state. Spatial data of water resource have been used for the extraction of tabular data of water resources as they are more understandable and useful in better utilization of water resources in agriculture.

## MATERIALSANDMETHODS

Various spatial data and digitized maps related to available water bodies of all districts of Bihar have been collected from different organizations and institutes. These data were arranged in the required format. After that, data mining techniques have been applied to extract data from maps in tabular form, which is more understandable. SQL, GIS software (Arc GIS 9.1), ANN have been used for data mining from digitized maps. Data mining techniques also include Classification, Clustering, Association rules. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. In data analysis model learns to predict a class label from a set of training data which can then be used to predict discrete class labels on new samples. One of the major goals of the classification algorithm is to maximize the predictive accuracy obtained by the classification model. Data mining classification algorithms follow two different learning approaches: supervised learning, unsupervised. The different classification techniques for discovering knowledge are Bayesian Networks (BN), Rule-Based Classifiers, Decision Tree (DT), Nearest Neighbour(NN), Support Vector Machine (SVM), Artificial Neural Network (ANN),Rough Sets, Fuzzy Logic, Genetic Algorithms (Beniwal and Arora 2012).

### Data mining using Structured Query Language (SQL)

Data mining is not particular to one type of media or data; it is applicable to any kind of information repository. The most commonly used query language for relational database is SQL, which allows retrieval and manipulation of the data stored in the tables as well as the calculation average, sum, min, max and count. For example, an SQL query to select the farmers grouped by category (Land Holding group) would be: SELECT count (*) FROM Subsidies WHERE type=small farmer GROUP BY category. As described in this paper, an SQL query to select the number of water bodies having area greater than 5 hectare (ha) of Patna district may be SELECT count (*) FROM Patna WHERE area >'5 ha'. Patna is the name of the table having standing water resource data of Patna district. In this way, the required data can be retrieved. Data mining algorithms using relational databases is more adaptable than data mining algorithms specifically written for flat files since they can take advantage of the SQL could provide, such as predicting, comparing, detecting deviations, etc.

### OLE DB with SQL for data mining

OLE DB (Object Linking and Embedding Database) is an API (Application Programming Interface) allows to access data from different databases which may relational or non relational/ attribute.OLE DB as a higher-level replacement for, and successor to, ODBC (Open database connectivity). OLE DB can be used in SQL database for accessing required data from different databases as data mining process. OLE DB is conceptually divided into two parts viz. consumers and providers. The consumers are the applications that require access to the data, and the providers are the software components act as the interface and provide the data to the consumer.

### On Line Analytical Processing (OLAP)

OLAP is an approach is used for multi-dimensional analytical queries and extracting required data Data mining is a part of OLAP with application such as forecasting or prediction in agriculture. It gives an opportunity to view agriculture data from different points of view to better understand what that data means OLAP has been used extensively for analysis of data of water resource as well as crop, soil, etc.. OLAP technique has been used to mine the available standing water resource data.

### Artificial Neural Networks

Artificial Neural Network is an analytic technique formed on the basis of the assumed learning process in the human brain. The same way the human brain after learning process is capable of deducing assumptions based on earlier observations, neural networks after learning process are

capable of predicting change and events in the system. Neural networks are a group of connected input/output units in which every connection hasits weight. Back propagation algorithm of ANN is used in this research work for data mining of standing water resource data of all districts of Bihar. ANN provides more accurate data than the other methods, but data extracted by Relational technique using SQL is also good and relevant.SQL, ANN, GIS, etc. have been used for data mining in this research work and paper. These tabular data may be used for better utilization of water resources in agriculture to increase crop productivity in Bihar because the tabular data is more understandable and easy to use.

## RESULTS AND DISCUSSION

Digital maps/ spatial data were collected from different organizations/ institutes and some water resource maps have also been created according to the requirement. After that non-spatial/ tabular data of standing water bodies in their number and area have been extracted from spatial data of all districts of Bihar. For example, the spatial and non-spatial/tabular (extracted) data of three districts viz. Madhubani, East Champaran and Patna (Figs.1, 2 and 3) and (Table.1, 2 and3) have been shown but tabular data have been extracted for all districts of Bihar.

### District wise availability and area of standing water bodies

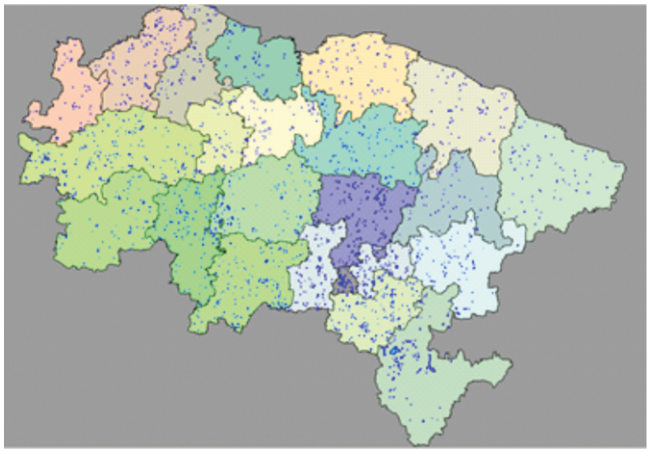

**Fig. 1**: Location of standing water bodies in Madhubani district of Bihar

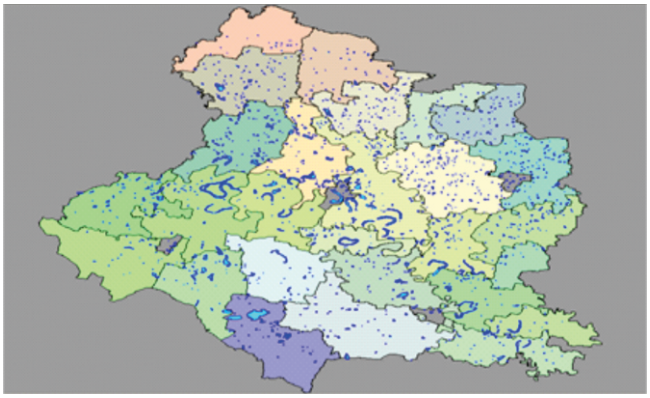

**Fig. 2:** Location of standing water bodies in East Champaran district of Bihar

**Table 2:** Categories of water bodies in Madhubani district of Bihar

| Name of District | Category of water body (Area-wise) (ha) | Number of water bodies | Total Area of water bodies (ha) |
|---|---|---|---|
| Madhubani | 0.5-10 | 2177 | 2155 |
| | 10-50 | 07 | 116.02 |
| | 50-100 | 01 | 83.96 |
| | 100-500 | 0 | 0 |
| | 500-1000 | 0 | 0 |
| | More than 1000 | 0 | 0 |
| | Total no. of water bodies | 2185 | 2355.42 |

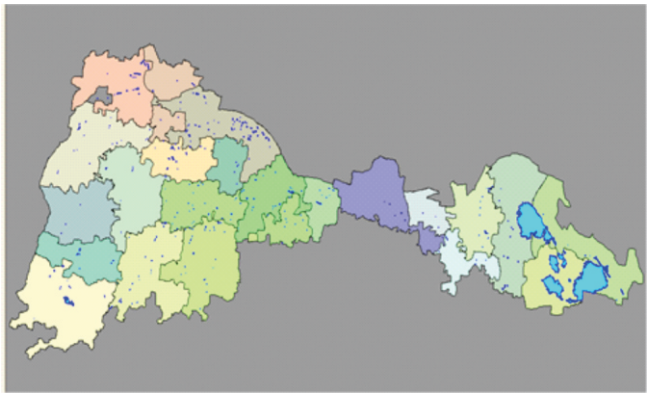

**Fig. 3**: Location of standing water bodies in Patna district of Bihar

**Table 2**: Categories of water bodies in East Champaran district of Bihar

| Name of District | Category of water body (Area-wise) (ha) | Number of water bodies | Total Area of water bodies (ha) |
|---|---|---|---|
| East Champaran | 0.5-10 | 1658 | 2318.86 |
| | 10-50 | 65 | 1376.44 |
| | 50-100 | 20 | 1320.95 |
| | 100-500 | 10 | 1736.11 |
| | 500-1000 | 0 | 0 |
| | More than 1000 | 0 | 0 |
| | Total no. of water bodies | 1753 | 6752.36 |

**Table 3**: Categories of water bodies in Patna district of Bihar

| Name of District | Category of water body (Area-wise) (ha) | Number of water bodies | Total Area of water bodies (ha) |
|---|---|---|---|
| Patna | 0.5-10 | 335 | 520.18 |
| | 10-50 | 09 | 136.08 |
| | 50-100 | 0 | 0 |
| | 100-500 | 03 | 681.78 |
| | 500-1000 | 01 | 936.03 |
| | More than 1000 | 02 | 6155.61 |
| | Total no. of water bodies | 350 | 8429.68 |

Among three districts, the maximum number of water bodies is available in Madhubani district, thereafter East Champaran and the least number of water bodies are in Patna district. But the maximum area of water bodies are present in Patna district, thereafter East Champaran and the Madhubani district has the least area of standing water bodies among three above mentioned districts of Bihar (Table 4 and Fig. 4). In Patna district, Mokama *tal* having less number of water bodies but the area of water bodies is very large because, after the

monsoon season, water becomes stagnant for few months in Mokama *tal* area because this area is low land area. After 3 to 4 months, pulse crops such as lentil, chickpea, pea and *tisi*, etc. have been cultivated. But 3-4 months, wetland remains idle. So there may be land-use planning for this *tal* area such as fish production, chestnut (Singhara) cultivation, etc. can be done in waterlogged period that will give farmers more income. This will uplift the livelihood of farmers. So these spatial and tabular data especially tabular data which is more understandable to farmers and a common person can be used for increasing the agricultural productivity and ultimately the income of farming community using these available excess standing water resources in these areas. So in this way, standing water resource data may be used to increase crop productivity for all districts of Bihar. In this way, these tabular data will be very useful for farmers to do better utilization of standing water bodies for increasing crop productivity.

**Table 4:** Number and area of standing water bodies of Madhubani, East Champaran and Patna districts of Bihar

| Districts | Number of SWB | Area of SWB (ha) |
|---|---|---|
| Madhubani | 2185 | 2355.42 |
| East Champaran | 1753 | 6752.36 |
| Patna | 350 | 8429.68 |

SWB: Standing Water Body

## CONCLUSION

It is true that there is a number of applications of data mining techniques in agriculture because a large amount of data are available from many resources in various formats such as relational or non-relational/ attribute or spreadsheet. The approach of integrating computer science with agriculture in forecasting and managing purpose.OLE DB connectivity works as API used for data mining technology that is very useful in accessing data from various sources. This technology is an excellent tool for doing an efficient implementation of data mining algorithms to achieve the complete database query and mining integration. These data are collected and stored in organized forms and integrated to form an information system. Data mining technique provides user-oriented access to new and hidden patterns in data, from which knowledge is generated which can help with decision making in agricultural organizations and these data may be used by researcher/ scientist for various agricultural purposes. So in this research paper, data mining techniques have been applied to extract useful non-spatial data in tabular form from spatial data/ maps. Relational properties of database using SQL statement, artificial neural network (ANN), GIS, etc. have been applied in data mining for getting useful district-wise data of standing water resources in their area and number of Bihar state that can be used for better utilization of water resources in agriculture to increase crop productivity and finally the income of farmers of this state.

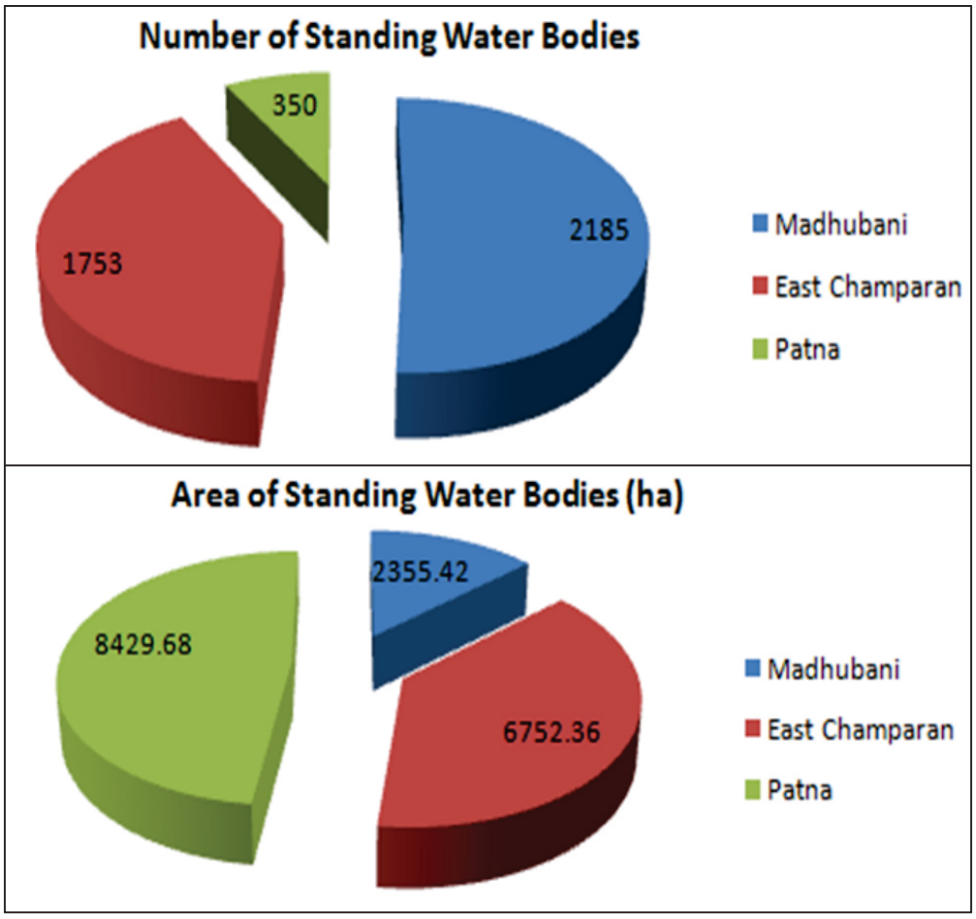


**Fig.4.** Number and area of standing water bodies of Madhubani, East Champaran and Patna districts of Bihar

**REFERENCES**

Abello J, Pardalos PM and Resende M. 2002. Handbook of Massive Data Sets, Kluwer Academic Publishers, New York, USA, 419-437

Ansari A and Ansari S. 2010. The concept of data mining, its applications,*Journal of Engineering and Sciences* **4(1):** 26-30.

Beniwal S and Arora J. 2012. Classification and feature selection techniques in data mining, *International Journal of Engineering Research & Technology* **1(6):** 1-6.

Cunningham S. J and Holmes G. 1999. Developing innovative applications in agriculture using data mining. In the *Proceedings of the Southeast Asia Regional Computer Confederation Conference*161-167.

Guang-xue Chen, LI Xiao-zhou, CHEN Qi-feng and LI Xiao-zhou. 2010. Clustering Algorithms for Area Geographical Entities in Spatial Data Mining, Seventh International Conference on Fuzzy Systems and Knowledge Discovery *IEEE*, 1630-1633.

Peng CH and Wen X. 1999. Recent applications of artificial neural networks in forest resource management: an overview. In: U. Corté and M.Sànche-Marrè (Eds.), *Environmental Decision Support Systems and Artificial Intelligence, AAAI Technical Reports WS-99-07*, AAAI Press, Menlo Park, CA, USA, 15-22.

Piatetsky-Shapiro G and Frawley W.J. 1991. *Knowledge Discovery in Databases*, AI Magazine, **13(3),**Menlo Park, CA, Association for the Advancement of Artificial Intelligence and The MIT Press, USA, 57-70.

Prasad Anup K, Chai Lim, Singh Ramesh P and Kafatos Menas. 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied earth observations and Geo information* **8(1):** 26-33.

Stastny J Konecny V and O Trenz. 2011. Agricultural prediction by means of neural network. *Agric. Econ.***57**(7): 356–361.

Xu Wei, Yong QIN and Houkuan Huang. 2003. The Application of Spatial Data Mining in Railway Geographic Information Systems. Proceedings of the IEEE International Conference on Intelligent Transportation Systems, 1467-1471.

**Citation:**Manibhushan and Ahmed A.2019. Spatial analysis of water resources data in selected districts of Bihar. *Journal of AgriSearch* **6**(3): 134-138