



Use of support vector machine and cellular automata methods to evaluate impact of irrigation project on LULC

Subhra Halder · Subhasish Das ·
Snehamanju Basu

Received: 13 March 2022 / Accepted: 25 July 2022 / Published online: 20 October 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract Land use and land cover (LULC) both define the earth's surface both anthropogenically and naturally. It helps maintain global balance but changes in land use create inequality. The LULC modification adversely affects physical parameters such as infiltration, groundwater recharge, surface runoff, ground temperature, and air quality. It is high time to monitor land use changes globally. Remote sensing and GIS techniques help to monitor these changes with a low budget and time. Various types of LULC classifiers have been invented to classify the LULC types. Maximum likelihood is a popular LULC classifier, but nowadays, support vector machine (SVM) classifier is gaining popularity because it provides a more accurate LULC than the maximum likelihood classifier. Therefore, in this study, the SVM classification technique has been applied to produce good accuracy

LULC maps. Using the SVM classifier, six LULC maps are produced from 1995 to 2020 for the Shali reservoir area in India which is a medium irrigation project irrigating ~3211 hectares of land per year. It plays an important role in the agricultural production of the region by providing irrigation water in monsoon and post-monsoon seasons. The impact of LULC change on the environment is also studied. The LULC forecast maps are also created using the cellular automata (CA) model and MOLUSCE plugin. Kappa coefficient and validation methods are used to validate the LULC and simulated maps. Both maps produce high accuracy with a kappa coefficient of 0.9. Secondary data, collected from the governmental gazette, such as population, crop production, and water level is also used to justify the results. The simulated map shows that 4% of agricultural land and built-up area may increase from 2020 to 2030. Overall, it has been proven that the SVM and CA models can produce accurate classified results.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10661-022-10588-6>.

S. Halder · S. Das (✉)
School of Water Resources Engineering, Jadavpur
University, Kolkata 700032, West Bengal, India
e-mail: subhasish.das@jadavpuruniversity.in

S. Halder
e-mail: subhrah.swr.rs@jadavpuruniversity.in

S. Basu
Lady Brabourne College, Kolkata 700017, West Bengal,
India
e-mail: drsnehamanju@gmail.com

Keywords LULC classifier · SVM classifier ·
Cellular automata model · MOLUSCE · Shali
reservoir

Introduction

The LULC classifies the uppermost layer of the earth, land cover describes the natural and anthropogenic features of the earth's surface, whereas land

use refers to the anthropological actions on the land surface. Both the terms are correlated; land use can modify the land cover and the modified land cover can transform the land use (Rawat & Kumar, 2015). LULC plays an active role in natural resource monitoring and conservation policies globally because of the adverse effects of LULC modifications nowadays. The changes in LULC affect physical parameters such as infiltration, groundwater recharge, surface runoff, land surface temperature, and air quality (Dale, 1997; Rounsevell & Recay, 2009; Dale et al., 2011; Dostdogru et al., 2020). The available terrestrial parts should be used rationally to maintain the ecological balance, so temporal change detection is considered mandatory. Currently, with the advancement of the remote sensing (RS) and geographical information system (GIS), the spatiotemporal change of LULC can be monitored easily on a low budget (Sinha et al., 2015). Since 1972, Landsat satellite data have been used for LULC classification. Various pixel-based measures are used to identify the classes (Lu & Weng, 2007) like the maximum likelihood classification (MLC), neural network classification, decision tree classifiers, support vector machine (SVM), k-means clustering, and fuzzy set classifier.

The MLC is the most popular classifier among all classification methods (Huang et al., 2002). The MLC is a parametric classifier, based on mean and covariance. It uses the probability theory to accomplish the classification. Here, the probability depends on the cell size, shape, and distance to the class centre. With a high probability value, it calculates the values of each raster cell and allocates the cell to the class (Bhatta, 2017). Although the MLC has some limitations, it needs representative pixel samples to get a normal distribution, which is not true for remotely sensed images (Dixon & Canade, 2008).

At present days, the SVM classifier is chosen for LULC classification to achieve good accuracy. The SVM was developed in the 1990s and is often considered one of the best classifiers. It is a non-parametric statistical technique that categorises data using hyperplane (James et al., 2013). Comparisons between the SVM and MLC have been published in several research papers. Based on the accuracy, SVM produces better accuracy than the MLC classifier (Karan & Samadder, 2016). The Global Positioning System (GPS) data have been used to verify the LULC classification. Verification discloses that the SVM method

is better than MLC, and the areal difference is comparatively less to ground data in SVM classification (Mondal et al., 2012). The SVM can produce good accuracy in LULC classification by generating hyperplane than MLC in high-resolution satellite imagery. The SVM is more advanced than MLC (Karan & Samadder, 2016). With the small number of pixels, SVM can assess better LULC classification than other classifiers (Kadavi & Lee, 2018). In comparison with Spectral Angle Mapper (SAM), which is a physically based spectral classification, the SVM shows precise and consistent accuracy in LULC classification, whereas SAM shows poor accuracy (Gopinath et al., 2019).

In the last few decades, significant development has been noticed in RS and GIS. Researchers are using RS and GIS in broader aspects of the LULC classification category (Ramachandra, 2016). Earlier in 2005, LULC classification was limited to identifying the present and past states of LULC (Aspinall, 2004). In recent times with the advancement of RS and GIS techniques, various models have been developed to predict future LULC patterns (Islam et al., 2018). Simulation techniques have also been introduced for remote sensing. With the help of simulation techniques, future land use prediction can be done easily (Xing et al., 2020).

The cellular automata (CA) model was developed by Ulam and von Neumann in the 1940s (Neumann & Burks, 1966) and is used for land use prediction. The five essential components of the CA model are cell area, cell positions, time phases, transition rules, and neighbourhood (Feng et al., 2019). The CA model is a separate dynamic system where the condition of each cell at time $t+1$ is governed by the condition of its adjoining cells, which is based on time and the pre-determined transition processes (Mohamed & Worku, 2020). The CA is a technique that can reproduce the sequential and spatial dynamics or growth of factors in two dimensions (Ye & Bai, 2007). Several CA-based packages have been developed to pretend land use like Dinamica EGO, UrbanSim, CLUES, SLEUTH, CA-Markov in UrbanCA, FLUS, and IDRISI. All the packages have their calibration, authentication, and valuation methods.

CA model requires a laborious calibration for better performance, which regulates the inner restrictions. The three assessment methods of CA are dataset assessment, which appraises the reliability

of the input dataset by inspecting the accurateness and outlines model errors with some useful visions to control the errors. Another method is a procedure valuation that assesses model path proficiency, CA transition guidelines, transition probability map, and model sensitivity and suggests procedures for correct modelling. The final method is result assessment that compares the modification among the actual and simulated outcomes by the graphical inspection, arithmetical analysis, spatial arrangement investigation, etc. (Tong & Feng, 2019).

A clear vision has come out from the research papers that the SVM provides a more accurate LULC classification than other classifier methods. Therefore, the objective of the study is to identify the temporal changes in LULC using the SVM classifier, and the second objective is to simulate future land use distributions using the CA model. For this purpose, the Shali reservoir area of India has been selected. Shali reservoir is a medium irrigation project that can irrigate about 3211 hectares of land per year, operating since 1996 in the Gangajalghati block of Bankura district in West Bengal of India. It performs a vital role in the agricultural productivity of this region by providing irrigational water for both the monsoon and post-monsoon seasons through the canal. Thus, it is important to monitor LULC changes in this region (Halder et al., 2020). This paper also explores the adverse effects of LULC on climate change.

Materials and methods

Study area

This study focuses on the chronological modification of LULC in the Shali reservoir area. The Shali reservoir is a medium irrigation project run by the Irrigation and Waterways Department of West Bengal, India. It is situated in the Gangajalghati Gram Panchayat of Bankura district in the State of West Bengal (Fig. 1). It provides irrigation water to the villages of Gangajalghati Gram Panchayat throughout the year. The total area of the study area (Gangajalghati Gram Panchayat) is around 44 km². The total area of the reservoir is around 1 km². The Shali reservoir can hold up to 361 ft (110.033 m) of water level, but when the water level rises to 357 ft (108.814 m), the excess water is discharged through the Shali reservoir

canal as a river elevator irrigation system for irrigational usage.

Here, the location of the Shali reservoir has been chosen as the study area because it has a great impact on LULC changes. According to the Meteorological Department in India, the Shali region receives less rainfall during the monsoon. The Shali area comprises rough and undulating terrain with a moderate to a high slope that enhances surface runoff and is covered with low red and laterite soil, which has low water holding capacity (Halder et al., 2020). These geophysical constraints are the main reason behind the construction of the Shali reservoir. Following the construction of the Shali reservoir, LULC pattern changes have been identified according to the irrigation information provided by the Irrigation and Waterways Department. Thus, the Shali reservoir plays a vital role in agricultural activities, and the local people are benefiting from this irrigation scheme. Therefore, the SVM classification method has been used to detect temporal changes in the land use pattern and the CA method for predicting the future land use pattern of the Shali region.

Gathered data

Remote sensing data

Landsat satellite images have been taken from the USGS earth explorer website (<https://earthexplorer.usgs.gov/>) to illustrate the periodical transformation in LULC. Images are taken from 1995 to 2020, and a 5-year gap has been maintained. All the images are taken for the month of January because this time, cloud cover remains very low in this region. Both the Landsat 5 TM and Landsat 8 OLI data have been incorporated in this study. All the images are geometrically corrected with the WGS 1984 UTM Zone 45 N coordinate system (Fig. 2).

Other data sources

The administrative boundary of the study area has been demarcated from open street maps (Open Street Map, <https://www.openstreetmap.org/>). Rainfall and water level data from the Shali reservoir were also used to signify the trend, and statistical techniques have been applied to identify the trend. Rainfall data were obtained from the Indian Meteorological

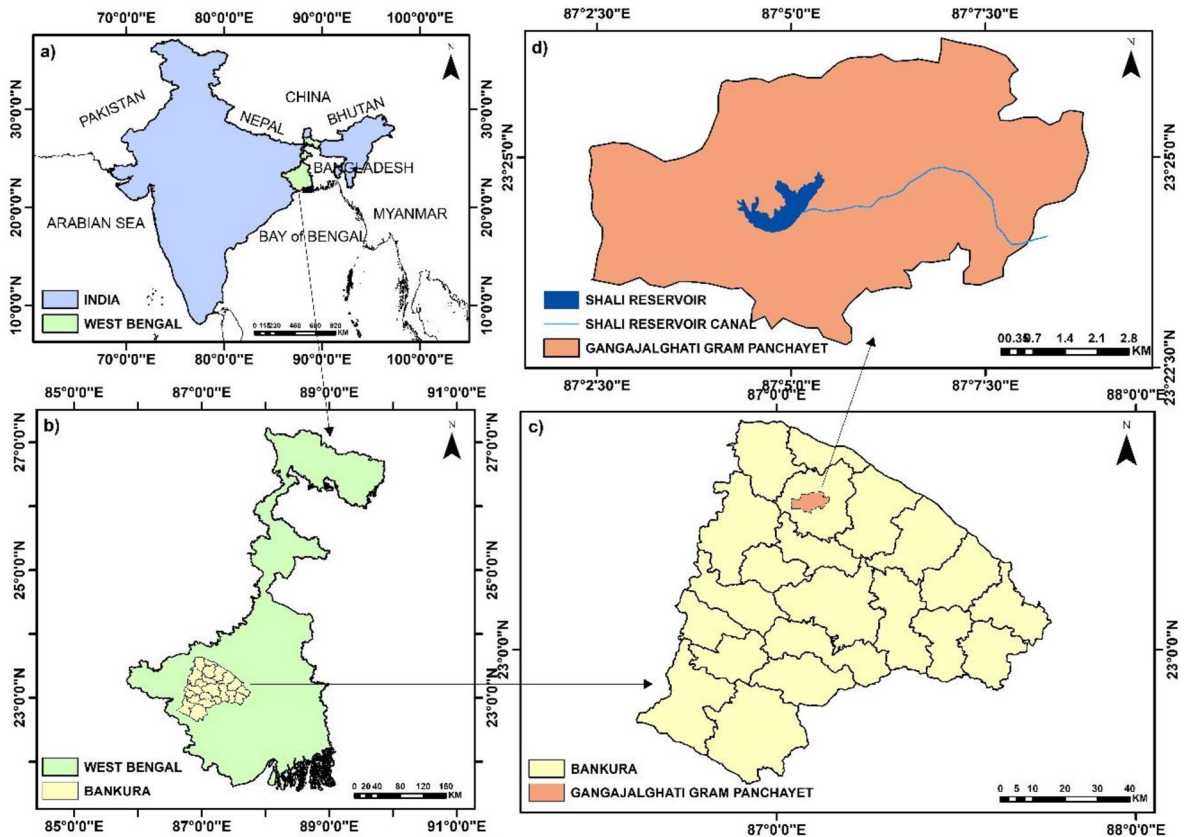


Fig. 1 Maps showing the location of the study area

Department (https://www.imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html) and water level data from the Irrigation and Waterways Department (<https://wbiwd.gov.in/>). In India, air quality index (AQI) is formed of major six pollutants such as particulate matter, sulphur dioxide, nitrogen dioxide, carbon monoxide, ozone, and ammonia. In this study, all these data were obtained from the NASA Giovanni data system (Acker et al., 2014). Then, AQI index has been calculated by using the AQI calculator (CPCB, 2022). The surface runoff data sets were taken from the TerraClimate Global data set (Abatzoglou et al., 2018; TerraClimate, 2020).

Methodology

The entire study has been divided into two segments. In the first segment, LULC classifications are performed using the SVM method from 1995 to 2020.

Change detection is also performed from 1995 to 2005 and from 2005 to 2020. In the second segment, the future prediction of land use is done using the CA approach. Here, MOLUSCE (Modules for Land Use Change Evaluation) a QGIS plugin, has been used to predict the future land use scenario. The plugin requires some specific input data sets such as LULC maps of 1995 and 2020, DEM, and road networks of the study area (Fig. 3). The entire study is based on (i) the SVM classifier and (ii) the CA model which are discussed below.

Support vector machine (SVM)

The SVM is a statistics-based method described by Cortes and Vapnik (1995). It is a non-parametric supervised classification method that can conduct high-resolution multiband satellite imagery. The main principle of the SVM is the formation of a

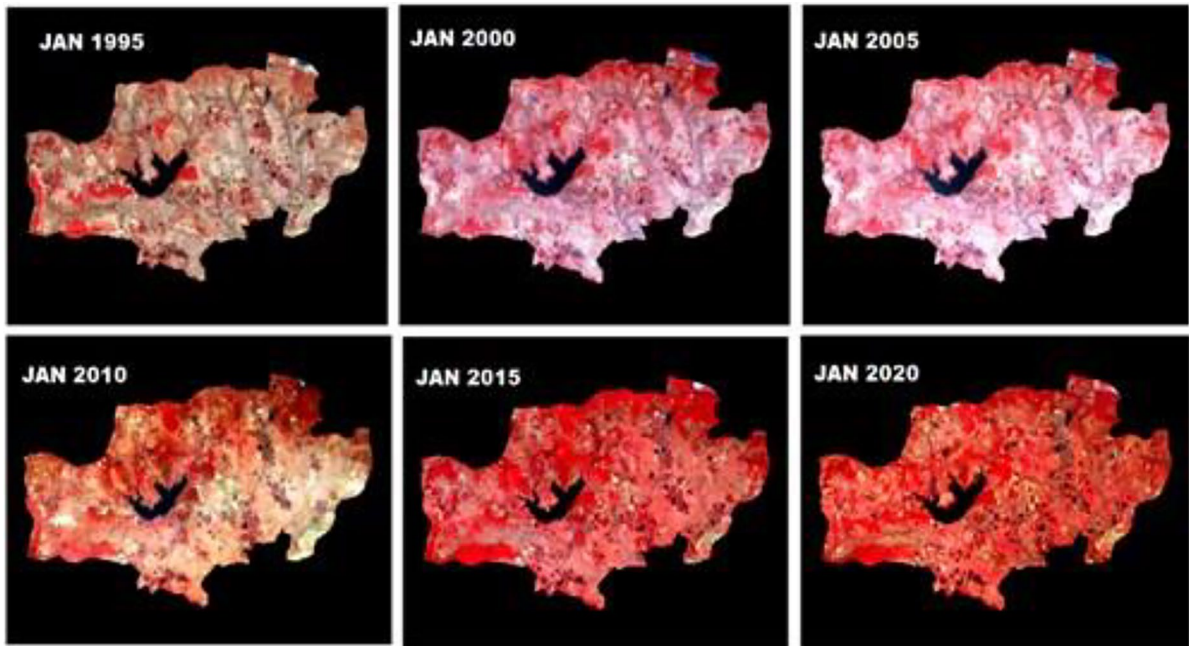


Fig. 2 Satellite imageries (FCC) used in the study

hyperplane; by generating a hyperplane, the data sets are classified into different classes. The SVM creates a concentrated boundary of separation along

the hyperplane between the classes. If the point is found above the hyperplane, it will be categorised as +1, and if not, it will be categorised as -1. The

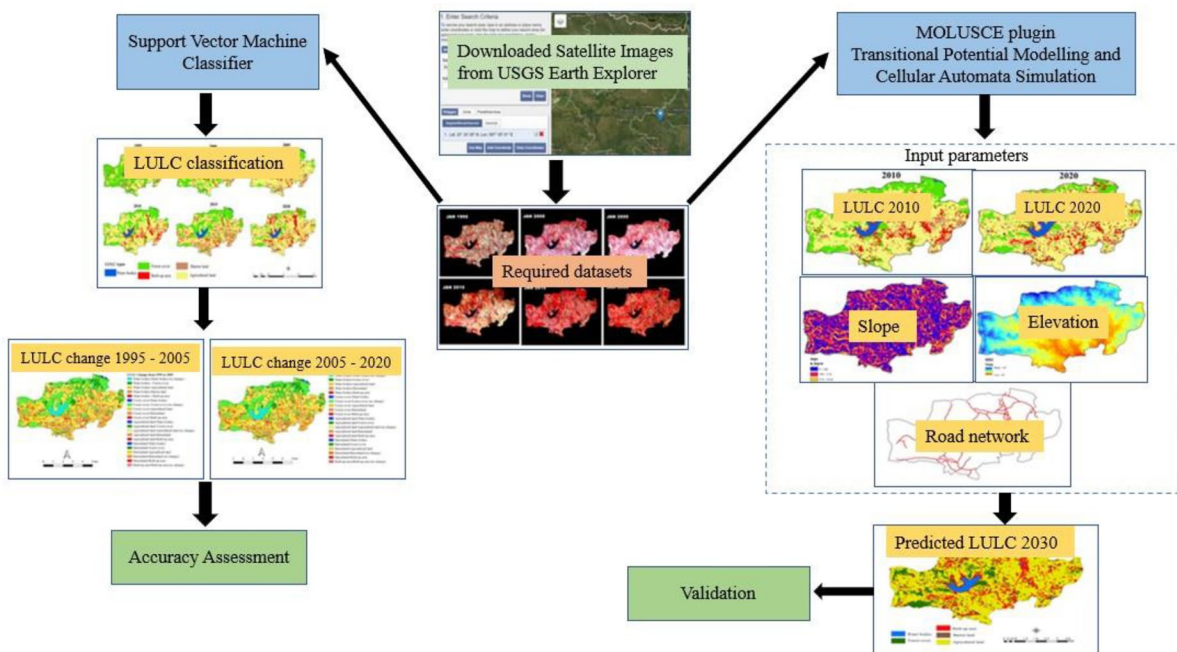


Fig. 3 Methodology adopted for this study

training points that adjoin the optimal hyperplane are named support vectors (Tehrany et al., 2014). The simplest way to train SVM is using linearly separable classes (Fig. 4) as described in Eqs. (1–2).

Set of labelled training patterns $(Y_1, X_1), (Y_2, X_2) \dots (Y_k, X_k) \dots Y_k \in [-1, +1]$ is called separable linearly if a vector W_k and a scalar B_k exist for the inequalities given in Eqs. (1–2).

$$W_k \times X_k + B_k \geq +1 \text{ for all } Y_k = +1, \text{ i.e. a member of class 1} \tag{1}$$

$$W_k \times X_k + B_k \leq -1 \text{ for all } Y_k = -1, \text{ i.e. a member of class 2} \tag{2}$$

These inequalities are valid for the entire training set (Y_k, X_k) . Here, k is quantity of samples signified as (Y_k, X_k) ; $x \in R^n$ is n -dimensional space; $y \in [-1, +1]$ = class label; W_k = perpendicular to the linear hyperplane; class 1 and class 2 represent -1 and $+1$.

One of the main advantages of SVM is that it produces less noise than other classifier techniques and can handle an uneven number of training data sets for each class. The purpose of SVM is to produce a model that forecasts the designation value of testing data from the training data attributes (Karan & Samadder, 2016). SVM has gained popularity in the remote sensing field because it can produce higher accuracy classification with the help of small training datasets than other traditional classification methods (Mountrakis et al., 2011).

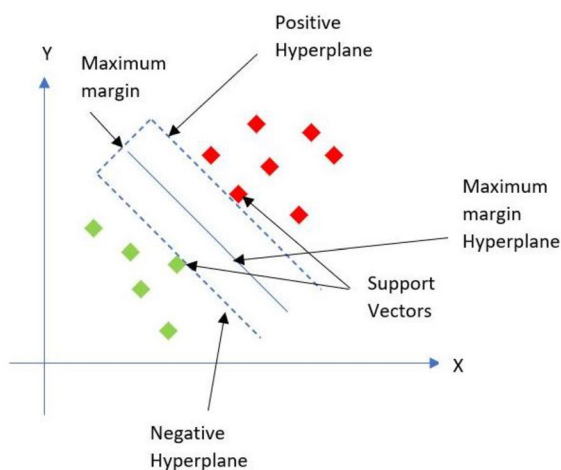


Fig. 4 Linear SVM method

Cellular automata (CA)

CA is a grid-based model with compact space computation power. It is a distinct dynamic system where the neighbouring cells control the condition of every cell. It is a process that can pretend the chronological and spatial dynamics in two dimensions. The CA model has been applied in universal and regional activities equally. The global model is unable to effectively express a specific location because it assumes that the variables are equal, whereas in the local analysis, considering the communication strength of the cell, there is a mutual evaluation among the distinctive situations of cells within a neighbourhood. The communication power of the cell reduces with the rise in the distance among the cells (Mohamed & Worku, 2020). The CA model can be stated as addressed by Muller and Middleton (1994) in Eq. (3).

$$S_{t+i} \geq f(S_t, N) \tag{3}$$

where S is a fixed, distinct position group of cells; N is the neighbouring cells; t and $t+i$ are distinctive moments, and f is the cell conversion rule of the regional space. In remote sensing, the CA model has been used to predict LULC classes (Lu et al., 2019).

Results and discussions

LULC classification using SVM method

From the literature reviews, it has been clear that the SVM classifier is considered one of the finest classification approaches of other types of classifiers. It provides better accuracy than the maximum likelihood classification method. In this study, the SVM method has been applied to get better accuracy classification. The SVM method was applied using the ArcMap 10.5 software. The LULC maps have been categorised into five classes, i.e., (i) water bodies, (ii) forest cover, (iii) agricultural land, (iv) barren land, and (v) built-up areas (Table 1).

The first step of the SVM classification is training the samples. Samples of different LULC cover types were collected from each year's satellite images. Then, to perform the classification, the collected samples were trained under the SVM classifier, and lastly, the raster images were classified using the trained samples.

Table 1 Delineation of LULC types

| LULC classes | Delineation |
|-------------------|--|
| Water bodies | Surface water bodies like lakes, ponds, reservoirs, tanks, canals, rivers |
| Forest cover | Land that is covered by natural forest covers like the evergreen forest, deciduous forest, open shrubs |
| Agricultural land | Crops like paddy, wheat, orchards, plantations, etc |
| Barren land | Exposed rocks, sandy areas, river banks, etc |
| Built-up | Industrial and commercial areas, residential areas, roads, etc |

The same process was performed for all the satellite images. The LULC maps have been prepared from 1995 to 2020 by maintaining a 5-year gap with five different classes: (i) water bodies, (ii) forest cover, (iii) agricultural land, (iv) barren land, and (v) built-up areas. The LULC maps show the temporal change in LULC categories of the research area (Fig. 5). Areal coverages and percentages of different LULC categories are shown in Table 2. The total area of the study area is around 43 km² with seven villages likely Keshria, Maraya, Bhairabpur, Mandi, Ganga-jalghati, Chhota Kumira, and Kenduadihi. Shali reservoir provides irrigational water to these villages throughout the year by Shali reservoir canal. Before the construction of the Shali reservoir, single-crop cultivation is noticed, but after the construction of the

reservoir, both pre-monsoon and post-monsoon crops are seen to be cultivated. According to the *Bankura District Statistical Handbooks (2010–2014)*, the major Kharif (monsoon) crop of this region is paddy and wheat, while potato and mustard are the main Rabi crops (winter) crops.

Figure 5 shows the transitional changes in the LULC of the research region. In 1995, about 3.96 km² area was under the water bodies. Then, in 2010, the area of water bodies was reduced to 1.76 km² but by 2020, the area of water bodies has increased to 2.2 km². In the case of the forest cover, a negative trend has been found. In 1995, about 24.2 km² area of this region was under the forest cover, and by 2020 around 4.84 km² area was under forest cover. Therefore, a major decline is noticed in the forest cover, although

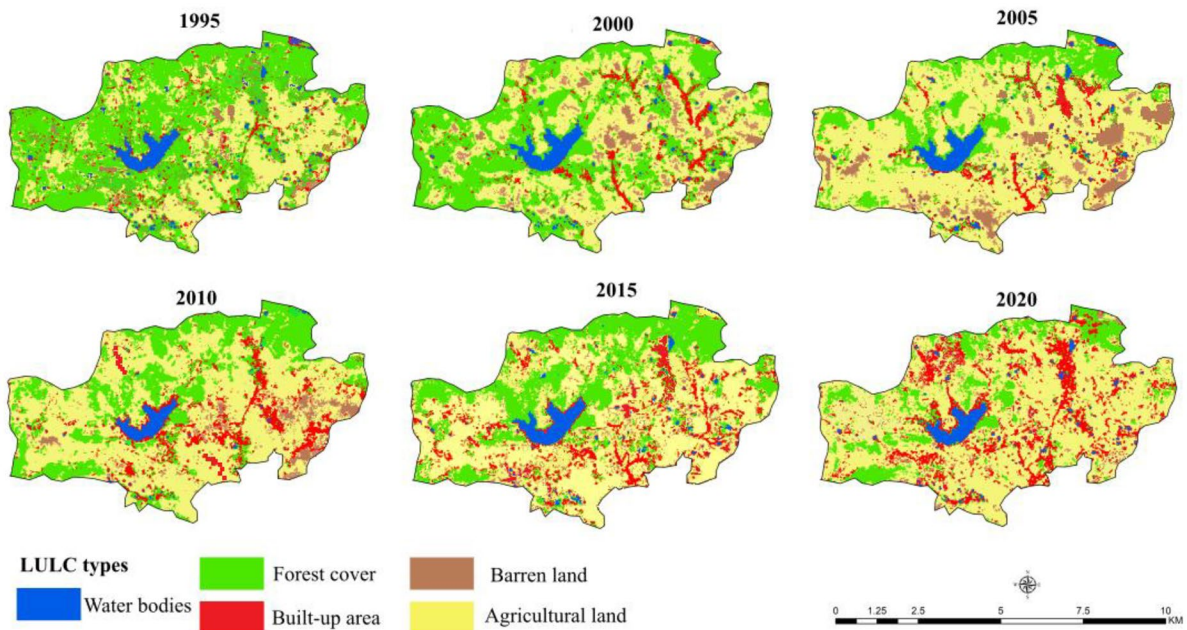


Fig. 5 LULC maps using SVM

Table 2 Areal coverages and percentages of different LULC classes

| LULC classes | Water bodies | | Forest cover | | Agricultural land | | Barren land | | Built-up area | |
|--------------|---------------|-----------|---------------|-----------|-------------------|-----------|---------------|-----------|---------------|-----------|
| | Area in sq.km | Area in % | Area in sq.km | Area in % | Area in sq.km | Area in % | Area in sq.km | Area in % | Area in sq.km | Area in % |
| Year | | | | | | | | | | |
| 1995 | 3.96 | 9 | 24.2 | 55 | 9.68 | 22 | 4.4 | 10 | 1.76 | 4 |
| 2000 | 3.96 | 9 | 15.4 | 35 | 14.96 | 34 | 7.04 | 16 | 2.64 | 6 |
| 2005 | 3.52 | 8 | 6.6 | 15 | 21.56 | 49 | 8.8 | 20 | 3.52 | 8 |
| 2010 | 1.76 | 4 | 9.24 | 21 | 22.00 | 50 | 5.28 | 12 | 5.72 | 13 |
| 2015 | 2.2 | 5 | 7.92 | 18 | 22.22 | 52 | 3.52 | 8 | 7.48 | 17 |
| 2020 | 2.2 | 5 | 4.84 | 11 | 24.64 | 56 | 2.64 | 6 | 9.68 | 22 |

agricultural land portions have increased from 1995 to 2020. In 1995 around 9.68 km² area was used for agricultural activities and increased through time. By 2020, about 24.64 km² area was subjected under agricultural activities. The portion of barren lands is also found to be declined over time. Barren lands have been converted into agricultural lands. In 2020, only 2.64 km² area has been found under the barren land category.

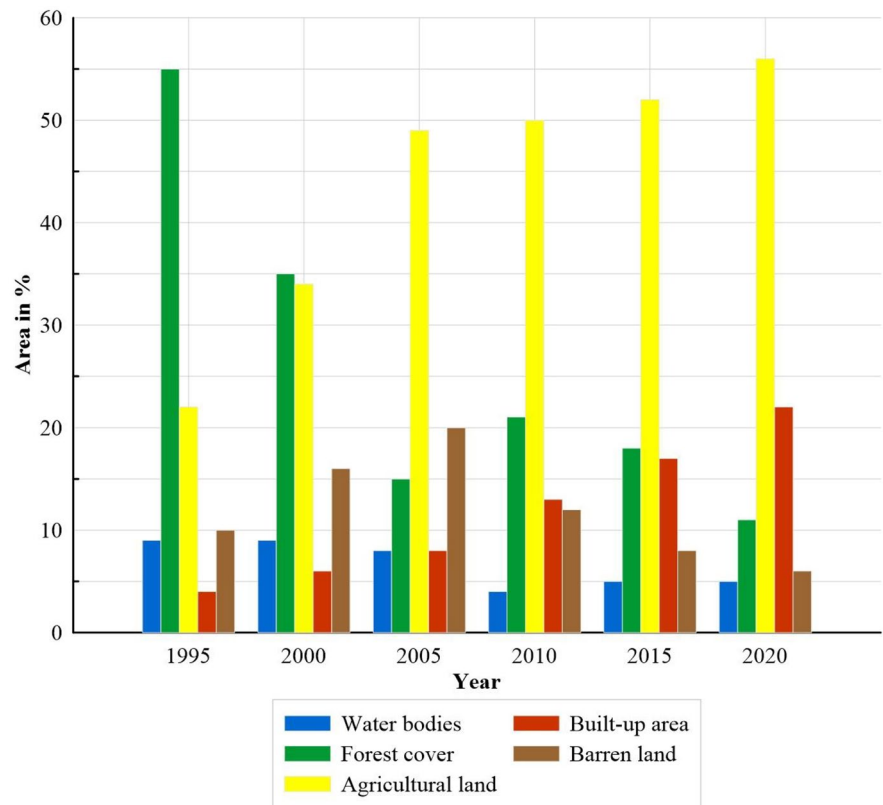
Therefore, after the creation of the Shali reservoir, agricultural productivity intensified in this region. Along the agricultural land, built-up area also increased through time. In 1995, around 1.76 km² was under the built-up category, but in 2020, the amount of built-up area has increased to 9.68 km². According to Indian Census (2011) population data, the population has increased in the Shali reservoir region over time. The population data justify the positive change in the built-up category of the LULC classification.

Therefore, LULC maps of the Shali area have revealed that significant changes have been observed in the LULC patterns since the creation of the reservoir. Prior to the construction of the Shali reservoir in 1995, an area of about 3.96 km² was under the water body category, but by 2020, it was reduced to 2.2 km². Prior to the construction of the reservoir, an area of about 24.2 km² was under forest cover, but after the construction of the reservoir, the amount of agricultural land increased rapidly, so rapid loss of forest cover has been observed in the study area. By 2020, only 4.84 km² area was under the forest cover category. Built-up areas also increased along with the agricultural lands after the creation of the Shali reservoir (Fig. 6).

Accuracy assessment of the classification

Here, supervised classification has been performed using the SVM technique to identify temporal changes in LULC of the Shali reservoir area. The five LULC categories used in this study are water bodies, forest cover, agricultural land, barren land, and built-up area. The LULC maps were created from 1995 to 2020 by maintaining a 5-year gap. An accuracy assessment was performed using the Google Earth software to validate the LULC results. Random 100 points were taken from the LULC maps to validate with the Google Earth images. The kappa coefficient, given in Eq. 4, has been applied to measure the accuracy (Cohen, 1960; Satya et al., 2020).

Fig. 6 Temporal modification in LULC



$$\text{kappa} = \frac{P_{0k} - P_{ek}}{1 - P_{ek}} \tag{4}$$

where P_{0k} signifies the proportion of actual agreements and P_{ek} signifies the proportion of expected agreements. The overall accuracy of the LULC maps for 1995, 2000, 2005, 2010, 2015, and 2020 was calculated 0.85, 0.90, 0.91, 0.94, 0.93, and 0.92, respectively.

Changes in LULC

Socio-economic development plays an important role in the transformation of LULC. In this paper, change detection has been done to identify the temporal change in LULC. Here, change detection is divided into two parts, i.e., from 1995 to 2005 and 2005 to 2020. Figure 7 displays the LULC change from 1995 to 2005 (Table 3). It shows the change in each category of LULC like water bodies turning into built-up areas and forest covers turning into the agricultural area and also shows the unchanged portions of the study area.

Figure 8 depicts the percentage change in LULC from 1995 to 2005 in the study area. It shows that water bodies have been reduced by around 1%, and forest cover has been decreased by about 40%. Whereas the amount of agricultural land, built-up, and barren land have been increased from 1995 to 2005, around 27% of agricultural land, 4% built-up, and about 10% barren land have increased in the Shali reservoir area. Table 3 shows the changes in LULC patterns in the study area. Figure 9 also displays the changes in LULC from 2005 to 2020 (Table 4). No change classes are shown in a comparatively light colour than change classes. The amount of change portion calculation was also made to identify the changes. From 2005 to 2020, a mixed change has been noticed. Agricultural lands and built-up areas have undergone a positive change from 2005 to 2020. Whereas, water bodies, forest cover and barren land showed negative change (Fig. 10).

Change detection analysis suggests that a positive change has occurred in agricultural land and built-up areas; around 34% of agricultural land and 18% of built-up area have been increased from 1995 to

Fig. 7 Changes in the detection of LULC from 1995 to 2005

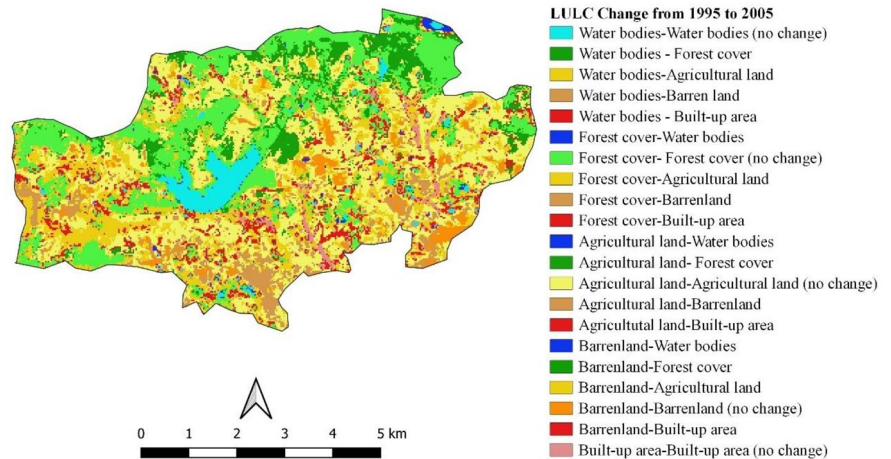
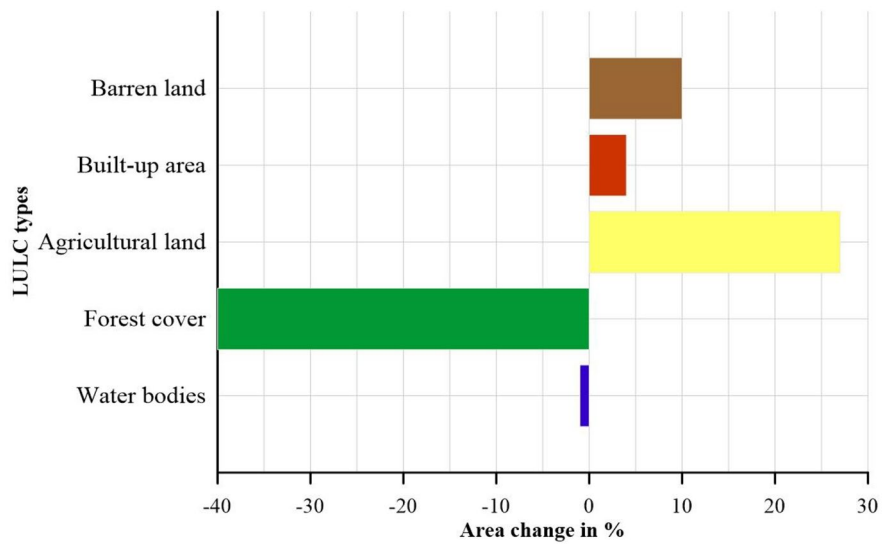


Table 3 LULC change matrix from 1995 to 2005

| Area in sq.km | | | | | | |
|-------------------|--------------|--------------|-------------------|-------------|----------------|--------------|
| LULC classes | Water bodies | Forest cover | Agricultural land | Barren land | Built-up areas | Total 2005 |
| Water bodies | 0.65 | 0.96 | 1.02 | 0.9 | 0 | 3.53 |
| Forest cover | 0.3 | 5 | 0.1 | 1.2 | 0 | 6.6 |
| Agricultural land | 0.15 | 13 | 7.58 | 0.83 | 0 | 21.56 |
| Barren land | 2.66 | 4.58 | 0.88 | 0.67 | 0 | 8.8 |
| Built up areas | 0.2 | 0.66 | 0.1 | 0.8 | 1.76 | 3.52 |
| Total 1995 | 3.96 | 24.2 | 9.68 | 4.4 | 1.76 | 44 |

Fig. 8 Percentage change in LULC from 1995 to 2005



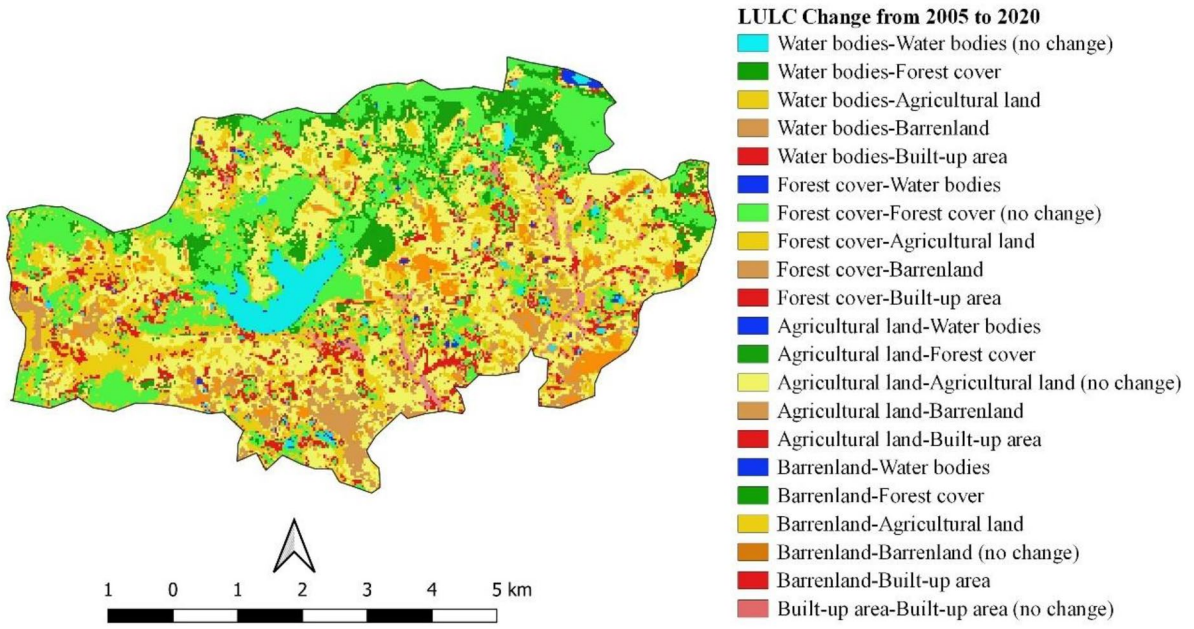
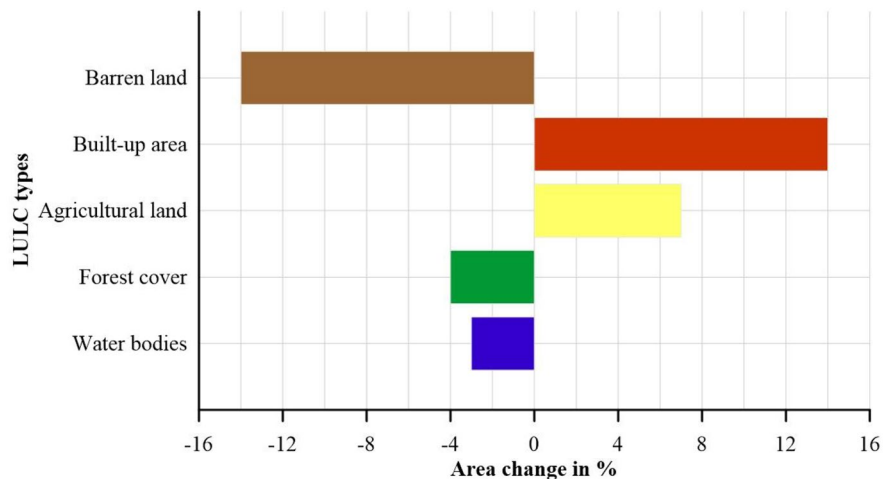


Fig. 9 Change detection of LULC from 2005 to 2020

Table 4 LULC change matrix from 2005 to 2020

| Area in sq.km | | | | | | |
|-------------------|--------------|--------------|-------------------|-------------|----------------|------------|
| LULC classes | Water bodies | Forest cover | Agricultural land | Barren land | Built-up areas | Total 2020 |
| Water bodies | 1 | 0.25 | 0.57 | 0.38 | 0 | 2.2 |
| Forest cover | 0.03 | 2.24 | 0.89 | 1.68 | 0 | 4.84 |
| Agricultural land | 0.4 | 1.8 | 17.74 | 4.7 | 0 | 24.64 |
| Barren land | 1.65 | 0.53 | 0.02 | 0.45 | 0 | 2.64 |
| Built-up areas | 0.45 | 1.78 | 2.34 | 1.59 | 3.52 | 9.68 |
| Total 2005 | 3.53 | 6.6 | 21.56 | 8.8 | 3.52 | 44 |

Fig. 10 Percentage change in LULC from 2005 to 2020



2020. A huge loss in forest cover has been observed in the study area, around 44% of forest cover has been reduced from 1995 to 2020.

The crop production area and irrigated area data were also referred from the District Statistical Handbooks (2010–2014) to validate the change analysis. After the creation of the Shali reservoir, both the crop production area and irrigated area have been increased. The MK (Mann–Kendall) Z values of crop production area and irrigated area are calculated at 5.42 and 7 which indicate an increasing trend. An increasing trend in crop production and irrigated areas is observed. Population data were also used to validate the increased rate of built-up category. The population data of the study region was compiled from the data obtained from the Census of India during the period 1991 to 2011. Population data shows a positive trend from 1991 to 2011, which justifies the

increase in built-up areas. So, it is proved that since the creation of the Shali reservoir, this region has progressed in agricultural activities, which helps increase built-up areas.

Problems due to LULC changes

The LULC classification of the Shali area reveals a significant change in forest cover in the study area. Around 44% of forest cover has been reduced from 1995 to 2020. This might affect the geophysical condition of the study area. To analyse the impact of LULC change on land surface temperature, surface runoff and air quality data (mean annual) of the study area have been analysed (Dale, 1997; Rounsevell & Reay, 2009; Dale et al., 2011; Dosdogru et al., 2020).

The study reveals that the LULC changes have created a negative impact on the environment of the

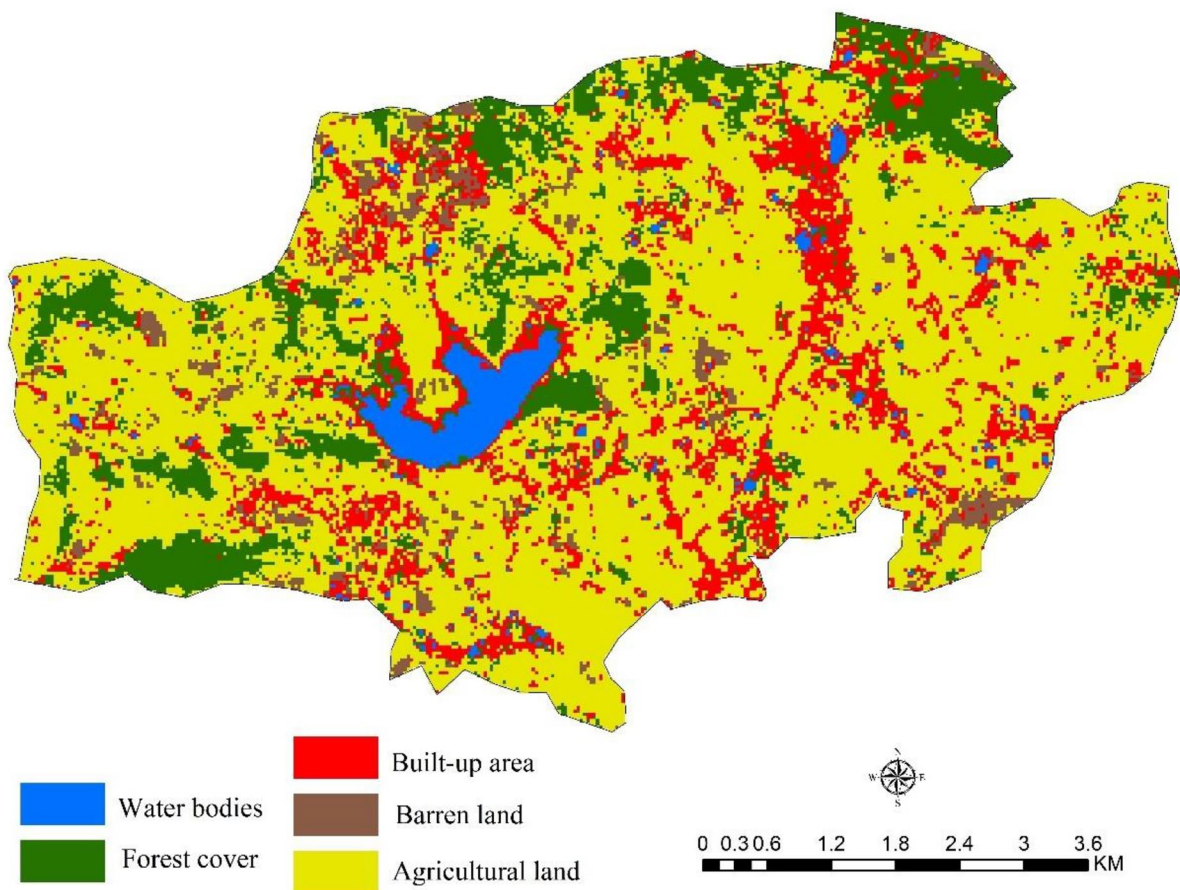


Fig. 11 Simulated LULC map for 2030

Shali reservoir area. The land surface temperature of the Shali reservoir area indicates an increasing trend (MK Z value 4.28) from 1995 to 2020. In 1995, the land surface temperature was around 24 °C, while in 2020 it increased to 31 °C. The surface runoff trend also shows a positive trend (MK Z value 4.19). The huge loss of forest cover enhanced the surface runoff in the study area. In 1995, the amount of surface runoff was around 9 mm, whereas in 2020, it increased to around 39 mm. The air quality is also affected by the change in LULC. In 1995, the air quality index of the Shali area was 56, which is under a satisfactory category, but in 2020, the air quality index was around 112, which is under the moderately polluted category.

Prediction of future land use

A prediction analysis has been conducted to forecast the future land use pattern of the Shali reservoir area. This prediction has been performed using the cellular automata (CA) model and artificial neural network (ANN) respectively which is done by MOLUSCE (Modules for Land Use Change Evaluation) plugin in QGIS software. It is an appropriate plugin to simulate LULC patterns. This plugin comprises three processes such as ANN, CA and validation.

The first step is the computation of the transition probabilities of the LULC maps by applying ANN. The second step is the calculation of transition potential maps, and the third step is the application of CA to stimulate the future LULC (Alam et al., 2021). The projection study is based on two kinds of variables: (a) dependent and (b) independent. LULC maps are considered independent variables, whereas slope, elevation, commercial area, residential area, educational

institutions, distance to road, distance to water bodies, etc. are considered dependent variables. The Euclidean distance function of ArcMap software was used to calculate the distance to road and water bodies. Slopes and elevations have been derived from SRTM DEM data.

In the simulation modelling process, after opening the MOLUSCE plugin, the first step is to input dependent and independent variables to estimate conversion matrices and varying probabilities. In the second step, the ANN model was used to predict the potential transition of LULC with 500 maximum iterations. The third step is to use the CA model to predict LULC and the final step is the validation of the predicted map (Kafy et al., 2021).

In this study, the MOLUSCE plugin is used to predict the LULC of 2030. Here, the LULC maps of 2010 and 2020 are considered independent input variables whereas slope, elevation, and distance to the road are considered dependent variables. Slope means the steepness of land, which can be measured in both degree and percentage. Elevation implies the height of a place from sea level and the distance to the road indicates areas close to the main road. Before the simulation of 2030, the LULC map validation process is ensued between the 2020 actual map and the predicted map.

The validation result shows around 95% correctness and an overall kappa value of 0.92 which proves good accuracy. After that simulated map was generated for the year 2030. The simulated map of 2030 is also classified into five different classes such as water bodies, forest cover, agricultural land, barren land, and built-up areas (Fig. 11 and Table 4).

According to the 2030 simulation map of the study area, water bodies occupy about 2% area, around 8%

Fig. 12 Areal coverage of LULC for 2030

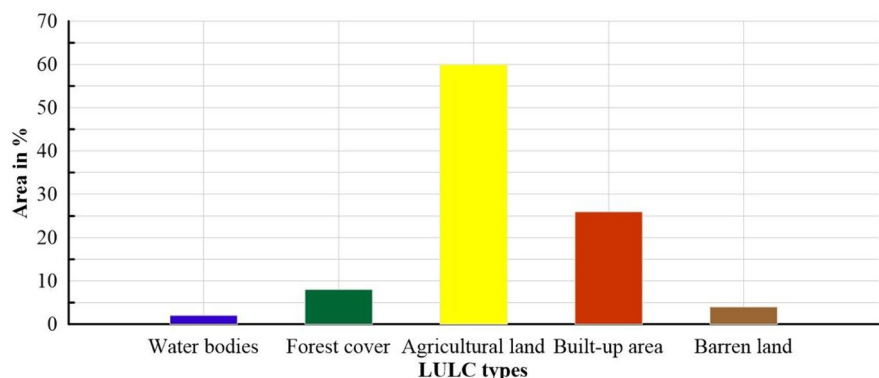
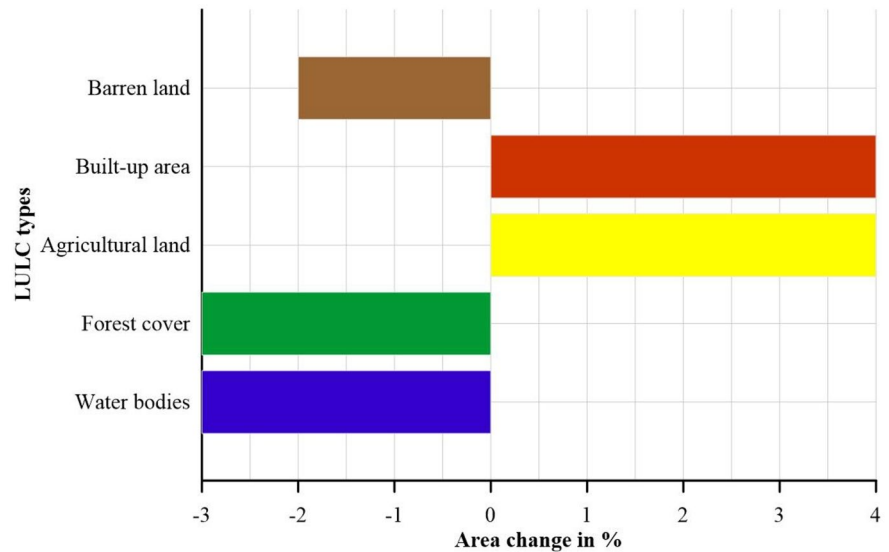


Fig. 13 Percentage changes in LULC from 2020 to 2030



area is under forest cover, agricultural land occupies about 60% area, about 4% area is covered with barren land, and 26% area is covered by built-up areas (Fig. 12). The change analysis has been done from 2020 to 2030. The result shows positive changes in the agricultural lands and built-up areas from 2020 to 2030 (Fig. 13 and Table 5).

Around 4% of agricultural land and 4% of built-up area have been increased, whereas water bodies, forest cover, and barren land have been reduced. Consequently, the proportion of agricultural land has increased after the creation of the Shali water reservoir, although water level data of the Shali reservoir have been included to justify the simulated result. Two forecast diagrams have been prepared using a time series analysis of the rainfall and water level data for the years 1995 to 2020. The Mann–Kendall test was also performed to identify the trends of the rainfall and water level data from 1995 to 2020. Time

series forecasting is a statistical method that uses time series data to make the prediction.

The time series analysis with the linear forecasting method has been used to forecast the water level data of the Shali reservoir, which shows that the water level will be increased in the future. In 1995 the water level of Shali reservoir was around 107 m, while in 2030 the water level will be increased to around 111 m. The MK (Mann–Kendall) test, the Z statistics is found to be 4.50, which means a positive trend in water level from 1995 to 2020.

The water level of the Shali reservoir depends on rainfall. Here, time series analysis and MK test have been applied to analyse the rainfall trend over the Shali reservoir area. It shows a positive trend in rainfall. The Z statistics in MK test is calculated 2.43, which means the amount of rainfall has increased through time. Thus, the rate of increase in agricultural land in 2030 is justified.

Table 5 Changes in LULC patterns in 2020 and 2030

| LULC classes | LULC in 2020 | | LULC in 2030 | | Change in area from 2020 to 2030 Area in % |
|-------------------|---------------|-----------|---------------|-----------|---|
| | Area in sq.km | Area in % | Area in sq.km | Area in % | |
| Water bodies | 2.2 | 5 | 0.88 | 2 | -3 |
| Forest cover | 4.84 | 11 | 3.52 | 8 | -3 |
| Agricultural land | 24.64 | 56 | 26.4 | 60 | 4 |
| Built-up | 9.68 | 22 | 11.44 | 26 | 4 |
| Barren land | 2.64 | 6 | 1.76 | 4 | -2 |

Conclusions

The purpose of this study was to focus on the use of support vector machine (SVM) and cellular automata (CA) methods to analyse temporary changes in the LULC and to predict the future LULC pattern of the Shali reservoir area which is a medium irrigation project.

From the LULC maps produced, a major change is seen in the LULC patterns. The reservoir plays an important role in agricultural development in the region from 1995 to 2020, with about 34% of agricultural land increased. The estimated Kappa coefficient, with an average value of 0.90, shows a good result. An analysis of the Mann–Kendall trend of irrigated and productive areas also confirms the results of the LULC classification. Therefore, it is recognised that the SVM produces an accurate classification. LULC modifications adversely affect physical parameters such as infiltration, groundwater recharging, surface runoff, ground temperature, and air quality. From 1995 to 2020, about 44% of forest covers were deforested and that accelerated soil temperature, surface run-off, and air quality of the influenced area.

The next goal was to predict future LULC patterns. The CA model is included in the simulation. The accompanying maps show that about 4% of agricultural land and built-up area are likely to grow from 2020 to 2030. The predicted LULC maps also confirm that agricultural development has taken place since the creation of the reservoir. The time series forecast method has been used in water level data to justify the simulated result. This result shows that water levels are increasing, which means that agricultural production also increases with water availability. Therefore, a combination of SVM and CA methods can be used to monitor the LULC pattern of a location, which will be beneficial in the near future.

Finally, the classification of the LULC reveals that after the construction of the dam, agricultural development has taken place, but at the same time, the changes are affecting the environment. Significant loss of forest cover has been observed during the study period. Time series of ground temperature, surface runoff, and air quality data show positive trends that negatively affect the environment. To address this situation, an annual afforestation program has been established by the government.

Acknowledgements The first author would also like to thank the University Grants Commission of India for funding a research fellowship for this work. The full cooperation from the Deputy Secretary to the Government of West Bengal, Irrigation & Waterways Department, and his entire team are also acknowledged for providing relevant field data related to Shali reservoir.

Data availability statement All data generated or analysed during this study are included in this published article.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Abatzoglou, T. J., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5, 170191. <https://doi.org/10.1038/sdata.2017.191>
- Acker, J., Soebiyanto, R., Kiang, R., & Kempler, S. (2014). Use of the NASA Giovanni data system for geospatial public health research: Example of weather-influenza connection. *International Journal of Geo-Information*, 3(4), 1372–1386. <https://doi.org/10.3390/ijgi3041372>
- Alam, N., Saha, S., Gupta, S., & Chakraborty, S. (2021). Prediction modelling of riverine landscape dynamics in the context of sustainable management of floodplain: A geospatial approach. *Annals of GIS*, 27(3), 299–314. <https://doi.org/10.1080/19475683.2020.1870558>
- Aspinall, R. (2004). Modeling land use change with generalized linear models—A multi-model analysis of change between 1860 and 2000 in Gallatin Valley. *Montana Journal of Environmental Management*, 72(1–2), 91–103. <https://doi.org/10.1016/j.jenvman.2004.02.009>
- Bhatta, B. (2017). Remote sensing and GIS. Third Edition. Oxford University Press.
- CPCB. (2022). Central Pollution Control Board, Ministry of Environment, Forest and Climate Change, Government of India. <https://cpcb.nic.in/>. (Accessed last on January 02, 2022).
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, XX(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Dale, V. (1997). The relationship between land-use change and climate change. *Ecological Applications*, 7(3), 753–769. [https://doi.org/10.1890/1051-61\(1997\)0070753](https://doi.org/10.1890/1051-61(1997)0070753)
- Dale, V. H., Efroymson, R. A., & Kline, K. L. (2011). The land use–climate change–Energy nexus. *Landscape Ecology*, 26, 755–773. <https://doi.org/10.1007/s10980-011-9606-2>
- District Statistical Handbooks. (2010–2014). Bankura 2010–11, Bankura 2013, Bankura 2014. Department of Planning and Statistics, Government of West Bengal. <http://>

- wbpspm.gov.in/publications/District%20Statistical%20Handbook. (Accessed last on December 27, 2021).
- Dixon, B., & Candade, N. (2008). Multispectral landuse classification using neural networks and support vector machines: One or the other, or both? *International Journal of Remote Sensing*, 29(4), 1185–1206. <https://doi.org/10.1080/01431160701294661>
- Dosdogru, F., Kalin, L., Wang, R., & Yen, H. (2020). Potential impacts of land use/cover and climate changes on ecologically relevant flows. *Journal of Hydrology*, 584, 124654. <https://doi.org/10.1016/j.jhydrol.2020.124654>
- Feng, Y. W., Wang, R., Tong, X., & Shafizadeh-Moghadam, H. (2019). How much can temporally stationary factors explain cellular automata-based simulations of past and future urban growth? *Computers, Environment and Urban Systems*, 76, 150–162. <https://doi.org/10.1016/j.compenvurbsys.2019.04.010>
- Gopinath, G., Sasidharan, N., & Surendran, U. (2019). Landuse classification of hyperspectral data by spectral angle mapper and support vector machine in humid tropical region of India. *Earth Science Informatics*, 13, 633–640. <https://doi.org/10.1007/s12145-019-00438-4>
- Halder, S., Das, S., & Basu, S. (2020). A review on the decadal irrigation system of Shali water reservoir. *Earth and Environmental Science*, 505, 012023. <https://doi.org/10.1088/1755-1315/505/1/012023>
- Huang, C., Davis, L. S., & Townshed, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), 725–749. <https://doi.org/10.1080/01431160110040323>
- <https://earthexplorer.usgs.gov/>. (Accessed last on November 17, 2021).
- <https://wbiwd.gov.in/>. (Accessed last on November 22, 2021).
- https://www.imdpune.gov.in/Clim_Pred_LRF_New/Gridded_Data_Download.html. (Accessed last on November 11, 2021).
- <https://www.openstreetmap.org/>. (Accessed last on November 11, 2021).
- Indian Census. (2011). Population data. India Ministry of Home Affairs, Government of India. <https://censusindia.gov.in/>. (Accessed last on November 11, 2021).
- Islam, K., Rahman, M. F., & Jashimuddin, M. (2018). Modeling land use change using cellular automata and artificial neural network: The case of Chunati Wildlife Sanctuary, Bangladesh. *Ecological Indicators*, 88, 439–453. <https://doi.org/10.1016/j.ecolind.2018.01.047>
- James, G. W., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with applications in R. *Springer, New York, NY*. <https://doi.org/10.1007/978-1-4614-7138-7>
- Kadavi, P. R., & Lee, C. (2018). Land cover classification analysis of volcanic island in Aleutian Arc using an artificial neural network (ANN) and a support vector machine (SVM) from Landsat imagery. *Geosciences Journal*, 22, 653–665. <https://doi.org/10.1007/s12303-018-0023-2>
- Kafy, A., Naim, M. N. H., Subramanyam, G., Faisal, A., Ahmed, N. U., Rakib, A. A., Kona, M. K., & Sattar, G. S. (2021). Cellular automata approach in dynamic modelling of land cover changes using Rapideye images in Dhaka, Bangladesh. *Environmental Challenges*, 4.
- Karan, S. K., & Samadder, S. R. (2016). Accuracy of land use change detection using support vector machine and maximum likelihood techniques for open-cast coal mining areas. *Environmental Monitoring and Assessment*, 188, 486. <https://doi.org/10.1007/s10661-016-5494-x>
- Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), 823–870. <https://doi.org/10.1080/01431160600746456>
- Lu, Y., Wu, P., Ma, X., & Li, X. (2019). Detection and prediction of land use/land cover change using spatiotemporal data fusion and the Cellular Automata–Markov model. *Environmental Monitoring and Assessment*, 191, 68. <https://doi.org/10.1007/s10661-019-7200-2>
- Mohamed, A., & Worku, H. (2020). Simulating urban land use and cover dynamics using cellular automata and Markov chain approach in Addis Ababa and the surrounding. *Urban Climate*, 31, 100545. <https://doi.org/10.1016/j.uclim.2019.100545>
- Mondal, A., Kundu, S., Chandniha, S., Shukla, R., & Mishra, P. (2012). Comparison of support vector machine and maximum likelihood classification technique using satellite imagery. *International Journal of Remote Sensing and GIS*, 1(2), 116–123.
- Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Muller, M. R., & Middleton, J. (1994). A Markov model of landuse change dynamics in the Niagara region, Ontario. *Canada. Landscape Ecology*, 9(2), 151–157. <https://doi.org/10.1007/BF00124382>
- Neumann, J. V., & Burks, A. W. (1966). Theory of self-reproducing automata. Edited and completed by Burks, A.W. University of Illinois Press, Urbana and London. <https://cba.mit.edu/events/03.11.ASE/docs/VonNeumann.pdf>. (Accessed last on January 02, 2022).
- Ramachandra, T. S. (2016). Stimulus of developmental projects to landscape dynamics in Uttara Kannada, Central Western Ghats. *The Egyptian Journal of Remote Sensing and Space Sciences*, 19(2), 175–193. <https://doi.org/10.1016/j.ejrs.2016.09.001>
- Rawat, J. S., & Kumar, M. (2015). Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. *The Egyptian Journal of Remote Sensing and Space Sciences*, 18(1), 77–84. <https://doi.org/10.1016/j.ejrs.2015.02.002>
- Rounsevell, M. D. A., & Recay, D. S. (2009). Land use and climate change in the UK. *Land Use Policy*, 26(1), S160–S169. <https://doi.org/10.1016/j.landusepol.2009.09.007>
- Satya, A. B., Sashi, M., & Deva, P. (2020). Future land use land cover scenario simulation using open source GIS for the city of Warangal, Telangana, India. *Applied Geomatics*, 12, 281–290. <https://doi.org/10.1007/s12518-020-00298-4>
- Sinha, S., Sharma, L. K., & Nathawat, M. S. (2015). Improved land-use/land-cover classification of semi-arid deciduous forest landscape using thermal remote sensing. *The Egyptian Journal of Remote Sensing and Space Sciences*, 18(2), 217–233. <https://doi.org/10.1016/j.ejrs.2015.09.005>

- TerraClimate. (2020). National Center for Atmospheric Research (NCAR) – Climate Data Guide. Boulder, United States. <https://climatedataguide.ucar.edu/>. (Accessed last on December 19, 2021).
- Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, 512, 332–343. <https://doi.org/10.1016/j.jhydrol.2014.03.008>
- Tong, X., & Feng, Y. (2019). A review of assessment methods for cellular automata models of land-use change and urban growth. *International Journal of Geographical Information Science*, 34(5), 866–898. <https://doi.org/10.1080/13658816.2019.1684499>
- Xing, W., Qian, Y., Guan, X., Yang, T., & Wu, H. (2020). A novel cellular automata model integrated with deep learning for dynamic spatio-temporal land use change simulation. *Computers and Geosciences*, 137, 104430. <https://doi.org/10.1016/j.cageo.2020.104430>
- Ye, B., & Bai, Z. (2007). Simulating land use cover changes of Nenjiang County based on CA-Markov model. Simulating Land Use/Cover Changes of Nenjiang County Based on CA-Markov Model. In: Li, D. (eds) Computer And Computing Technologies In Agriculture, Volume I. CCTA 2007. The International Federation for Information Processing, vol 258. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-77251-6_35

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.