Correlation Between Built-Up and Land Surface Temperature Using Sentinel-2 and Landsat-8 Images through Semi-Supervised Deep Learning Model for Efficient Land Surface Monitoring

Saurabh Srivastava [#1], Tasneem Ahmed [*2]

Advanced Computing Research Laboratory, Department of Computer Application,

Integral University, Lucknow, Uttar Pradesh, India- 226026

srsurabh@iul.ac.in, tasneemfca@iul.ac.in

Article Info
Page Number: 3000-3023
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract

Satellite images are a powerful tool of earth observation that are providing a different view of earth surface monitoring. Currently, more than 200 satellites are available in orbit, and some of them are operational which are producing different kinds of images including optical, SAR, and thermal images with different spatial, spectral, and temporal resolutions. Many researchers have developed various frameworks, models, and systems that are using machine learning and neural network techniques for Earth's surface monitoring such as agricultural monitoring, urban area monitoring, flood monitoring, soil monitoring, etc. Satellite image contains heterogeneous characteristics of big data so machine learning and neural network techniques are less capable to produce an efficient result. Therefore, to overcome this problem, a semi-supervised deep learning model is proposed that is capable to classify the images into four land cover classes such as urban, barren lands, water bodies, and agricultural fields, and also capable to measure the land surface temperature (LST). In this paper, Sentinel-2 and Landsat-8 optical satellite images of the summer season from 2017 to 2021 have been used to estimate the urban area and land surface temperature of Lucknow city situated in Uttar Pradesh, India. The Sentinel-2 image is a high-resolution satellite image so it has been used in urban monitoring and the Landsat-8 image have two Thermal-Infrared bands that are used to measure the temperature of the land surface. Further, estimated urban area and LST have been used to find out the correlation between them, and are also used to conduct the trend analysis for the 10 years (2022 to 2031) using the linear trend analysis technique. The LST has been increasing in the taken period, according to trend analyses. It is also observed that LST is heavily influenced by the growth and development of built-up areas as well as the alteration of the thermal characteristics of the urban core and barren lands. In the future, this type of study may be very useful to construct the plan urban areas, which may be very helpful to control the temperature rise.

Keywords: -SVM; ANN; ML; PCA; Urban Area; Land Surface Temperature

Introduction

Urbanization is a widespread and intricate phenomenon that simultaneously involves people moving from rural to urban areas, changing from agriculture to manufacturing and service industries, and converting naturally occurring land into artificial urban environments [1]. The process of urbanization started with the establishment of towns and cities, which grew in size to become metropolitan and urban agglomerations. The process of urbanization is complex and dynamic, it entails alterations to the built environment's physical and functional components, hastening the shift from landscape to urban [2]. Unplanned and uncontrolled rapid growth has had serious detrimental consequences for city people and the environment [3]. The urban agglomerations in India such as Mumbai, Delhi, Kolkata, Bangalore, Chennai, and Hyderabad have risen very fast, and India is predicted to have the world's biggest concentration of urban agglomerations within a few years [2]. Rapid urbanization includes a variety of environmental challenges such as urban heat, rising energy consumption, pollutant emissions, environmental degradation, and urban climate change [4]. Urban areas have much more complicated scattering mechanisms compared to natural areas due to the high degree of landscape variability, complex combinations of natural and artificial elements, forms and sizes of various materials and objects, and complex combinations of natural and artificial elements [5]. Accurate and timely data is required to understand urbanization processes, urban land-cover changes, and tackling environmental issues using quantitative urban area descriptions and spatial distributions [4]. Managing urbanization has become a difficult task in India, and urbanization can be described as a change of land that occurs mainly as a result of rural-to-urban migration [2]. India's urbanization is neither a unique phenomenon, nor it is exclusive. It is similar to other regions of the world. India's urbanization has progressed, and provide economic transformation as a result [3]. Most suburban development lacks planning, which has led to greater transportation costs for suburban residents in terms of money, time, and inconvenience, high public sector expenses, unfavorable land use patterns, and an insufficient amount of open spaces, recreational facilities, and other facilities [3]. LST affects the radiation balance, turbulent heat fluxes, evapotranspiration, and other critical climatic factors in urban context. Thermal remote sensing is helpful in the analysis of these findings. The extent and intensity of the surface urban heat are frequently measured using satellite thermal infrared (TIR) data because of the well-known benefit of the satellite geographic coverage [6].

Satellite images are an important tool for tracking changes in land cover caused by temporal dynamics and changes in the earth's surface [7]. Satellite image contains rich information about the Earth's surface, and play an important role in Earth surface monitoring, interpretation, and analysis. Quantitative and qualitative information is provided by satellite images that help to minimize the complexity of fieldwork and study time. Satellite images are acquired at regular intervals using the various optical, microwave, and thermal sensors [8]. Information on both artificial and natural resources can be acquired by satellite images which have been used for a number of purposes including time series analysis, change detection, vegetation monitoring, and mapping of land use and land cover [9], crop monitoring, soil mapping characteristics, forest cover mapping differences, land cover change detection, natural disaster assessment, water resource applications, and wetland mapping, etc. [10]. Currently, there are a large number of satellites in orbit acquiring the images of the Earth surface or other planets on a daily, weekly, or monthly basis using a variety of sensors to collect multi-spatial and multi-temporal data [9]. The United States and the European Union are providing an unparalleled amount of satellite images with improved geographical, spectral, and temporal resolution due to open data rules [11]. The high and mid-resolution optical satellites such as Sentinel-2 and Landsat-8 image can provide a better result for earth surface interpretation and analysis. Sentinel is a European radar imaging satellite that was put into orbit by the European Space Agency (ESA). It offers images in all kinds of lighting and weather. Sentinel is envisioned as filling the void left by the soon-to-be-phased-out existing order of earth observation missions. The European Space Agency (ESA) was launched Sentinel-2 on June 23, 2015 [12], and the main objective of sentinel 2 is to provide data for risk management, mapping of land use and land cover, change detection, detecting natural disasters, and water management. The USGS EROS Centre developed Landsat-8, which contains features an operational land imager (OLI) and thermal infrared sensor (TIRS) with 11 multi-spectral bands [12]. The National Aeronautics and Space Administration operate Landsat-8, and it is a mid-revolutionary optical satellite image. The Landsat series has been used for many different purposes, including monitoring of land and agriculture, geology, forestry, regional planning, classifying land use and land cover, and more [12].

The mapping of land cover and land cover change over large areas is made possible by satellite images which enable for consistent and reliable monitoring of the Earth's surface [13]. Urban planning, resource management, and mapping all depend on an understanding of the geographical location and growth of urban areas [14]. In order to achieve a substantial scientific purpose, land cover mapping and monitoring are conducted such as the data produced can be used to support environmental and atmospheric models, decision-making processes, etc. [15]. Maps of land use and land cover are essential for environmental modeling and the management of natural resources like water, as well as for understanding how people and the environment interact [16]. Monitoring land use change in urban areas can be useful to making decisions about resource management and urban planning. The satellite image is a more affordable alternative to ground-based surveys for land-use cover mapping and change analysis. Time series satellite images are used to examine the temporal dynamics of urban characteristics or processes [17]. The monitoring of land-use changes can be done rapidly and easily using satellite image classification, and the produced result provides efficient data support for enhancing land comprehensive usage [18]. Land use and land cover maps are made by classifying images that may be typically aerial photography or satellite images [15]. Satellite image has become widely available in the previous decade. Satellite images have recently become popular for classification and change detection in urban areas [10]. Satellite images are exceptionally helpful for a variety of purposes including monitoring of the Earth's surface and predicting and mitigating disasters. The most crucial method for collecting information on numerous monitoring activities on the surface of the Earth is satellite image classification [10]. Various image classifiers are available that have been used with satellite images globally to obtain information of the Earth’s activity. Traditional classification techniques are mostly probability based such as Bayesian decision theory with certain assumptions. Modern technologies are based on neural networks that produce efficient classification results [10]. Due to the great spatial and spectral resolution of information accessible from Earth observations, an object-based image analysis approach attracts attention when evaluating satellite images [19].

The classification of satellite images helps in collecting information regarding numerous monitoring activities on the surface of the earth which are difficult to overstate [10]. The process of selecting a suitable classifier for satellite image classification is a challenging task, and it plays an important role because the suitable classifier can produce an efficient and accurate result. The classification system must be broad, instructive, and removable [10]. Random Forest, Decision Tree, and Logistic Regression are the machine learning algorithms that have been used to classify satellite images [11]. There are numerous techniques and strategies that can be used to classify satellite images but majorly classification techniques can be categorized into three categories such as automated, manual, and hybrid [8]. Automated satellite image classification algorithms systematically organize pixels into meaningful groups across the whole satellite image. This category includes the vast majority of classification methods. The automated satellite image classifiers are automatically classifying satellite images and it can be divided into two groups such as supervised and unsupervised [8]. In supervised satellite image classification, the training sample is the most significant aspect, because the accuracy of the approaches is strongly dependent on the training samples. There are two types of training samples, first for classifying the image and the other for ensuring classification correctness [8]. Utilizing people's prior knowledge can increase the supervised classification algorithms' accuracy. The traditional supervised classification approaches compute the classes/cluster using the average value of all samples from the sample region [18]. The unsupervised classification is non-parametric technique, and in unsupervised classification, no prior knowledge is required to classify the images. It divides the pixels into classes based on their similarity without any training data [20]. As unsupervised classification methods, clustering techniques group the pixels in satellite images into unlabeled groups or clusters. After that, the analyst gives each cluster the appropriate labels and produces a well-classified satellite image. ISODATA and K-Means are the most common unsupervised satellite image classification algorithms [8]. Unsupervised classification approaches are time and money efficient and do not require prior expertise in the research area. In practice, however, the classification accuracy does not satisfy the standards due to a lack of information on ground calibration [18]. Manual image classification techniques are reliable, effective, and efficient. The analyst must be conversant with the area covered by the satellite image when using manual approaches. The classification's efficiency and correctness are determined by the analyst's knowledge and familiarity with the topic of research [8]. The advantages of automatic and manual methods are combined in hybrid satellite image categorization systems. In a hybrid approach, automatic satellite image classification methods are used for initial classification, and then manual approaches are used to refine and repair errors [8].

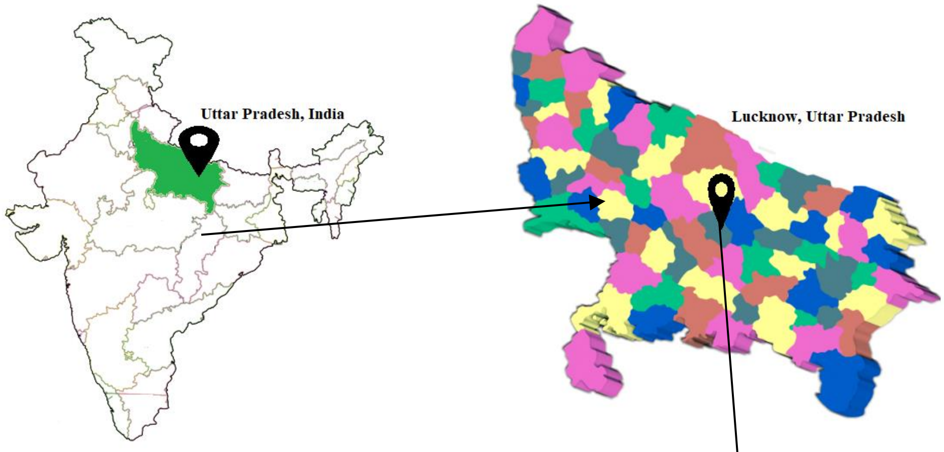
The advancement in technology of satellite images for Earth observation has become a significant landmark because of its accuracy in different application areas. However, it is required to employ satellite images for assessments of the Earth's surface due to the difficulties of the task and the limitations of sensor resolutions at high altitudes which requires a lot of time and effort in the field [21]. The capability of land cover monitoring is currently being enhanced following the availability of satellite images by utilizing dense time series, and cost-effective classification algorithms [22]. Modern computer technology and sensor system are capable to acquire satellite images that have good spatial, spectral, and temporal resolution [23]. The multispectral data analysis method can be used as an optical quantitative method for achieving fine spectral resolution in satellite images [23]. Satellite image technology can be used to monitor the dynamics of land use and land change information quickly and accurately, and improving image classification accuracy is a key aspect of monitoring changes in a category using remote sensing data [18].

In this study, a semi-supervised deep learning model has been developed to find out the correlation between the urban areas and land surface temperature. Sentinel-2 images have been used in this study for urban area estimation and the Landsat-8 images have been used to estimate land surface temperature. The estimated values of this study have been used to calculate the correlation between them, and also used to predict the trend of the urban area and land surface temperature for the next 10 years (2022 to 2031) using the linear trend analysis.

Study Area and Satellite Images Used

Study Area

Lucknow city the capital of Uttar Pradesh and its nearby areas are considered as the study area and it is located between Latitude 26.846694◦ and Longitude 80.946166◦. As a thriving center of North Indian culture and the arts and as the seat of the Nawabs' authority in the 18th and 19th centuries, Lucknow has always been a multiethnic metropolis. It continues to be a key hub for state administration, government, education, business, finance, pharmaceuticals, technology, design, culture, tourism, music, and poetry. The geographical location of the study area and Google Earth's image are shown in Figure 1(a) and 1(b).



(a) Geographical Location of Study Area

(b) Google Earth Image

Figure 1. Study area images, (a) Geographical Location of Study Area, and (b) Google Earth Image

Satellite Images Used

Sentinel-2

Sentinel-2 has a multispectral imager (MSI) with 13 bands [15]. In this study, images of the summer season from 2017 to 2021 of Lucknow city have been used for urban area estimation, and the details of the used images are mentioned in Table 1.

Table 1. Detailed of Sentinel-2 Satellite Images

Acquisition Date	Image Id	Sensor Name	Acquisition ID
May 14, 2017	Img1	Sentinel-2	L1C_T44RMQ_A009883_20170514T051753
May 29, 2018	Img2	Sentinel-2	L1C_T44RMQ_A015317_20180529T051027
May 24, 2019	Img3	Sentinel-2	L1C_T44RMQ_A020465_20190524T051014

May 8, 2020.	Img4	Sentinel-2	L1C_T44RMQ_A025470_20200508T050657
May 23, 2021	Img5	Sentinel-2	L1C_T44RMQ_A030904_20210523T050924

Landsat-8

The Landsat-8 OLI sensor has nine reflective bands with a spatial resolution of 30 meters, while the Landsat-8 TIRS sensor has two TIR bands (Band 10 and Band 11). Eight bands including visible, near-infrared, and shortwave infrared are used by OLI to collect data with a 30 meter spatial resolution. A panchromatic band with a 15 meter spatial resolution is also included. TIRS used two bands to sense TIR radiance with a spatial resolution of 100 meters, in the atmospheric window between 10 and 12 meters [23]. In this study, images of the summer season from 2017 to 2021 of the study area have been used for land surface temperature (LST) measurement, and the details of the used images are mentioned in Table 2.

Table 2. Detailed of Landsat-8 Satellite Images

Acquisition Date	Image Id	Sensor Name	Acquisition ID
May 10, 2017	Img6	Landsat-8	LC08_L1TP_144041_20170510_20170516_01_T1
May 13, 2018	Img7	Landsat-8	LC08_L1TP_144041_20180513_20180517_01_T1
June 1, 2019	Img8	Landsat-8	LC08_L1TP_144041_20190601_20190605_01_T1
May 18, 2020	Img9	Landsat-8	LC08_L1TP_144041_20200518_20200527_01_T1
June 6, 2021	Img10	Landsat-8	LC08_L1TP_144041_20210606_20210614_01_T1

Theoretical Background

Pre-processing of Sentinel-2 and Landsat – 8 OLI Images

Processing satellite images is a crucial step in understanding and analyzing satellite images. There are numerous satellite image repositories that offer satellite images for analysis and interpretation. The raw satellite image is not able to produce an efficient result because the satellite image contains multiband with different resolutions so the combination of bands cannot be used together for interpretation and analysis of the image. In this study, the Sentinel-2 and Landsat-8 images have been used for urban area monitoring and land surface temperature measurement, and the process of preprocessing is shown in Figure 2.

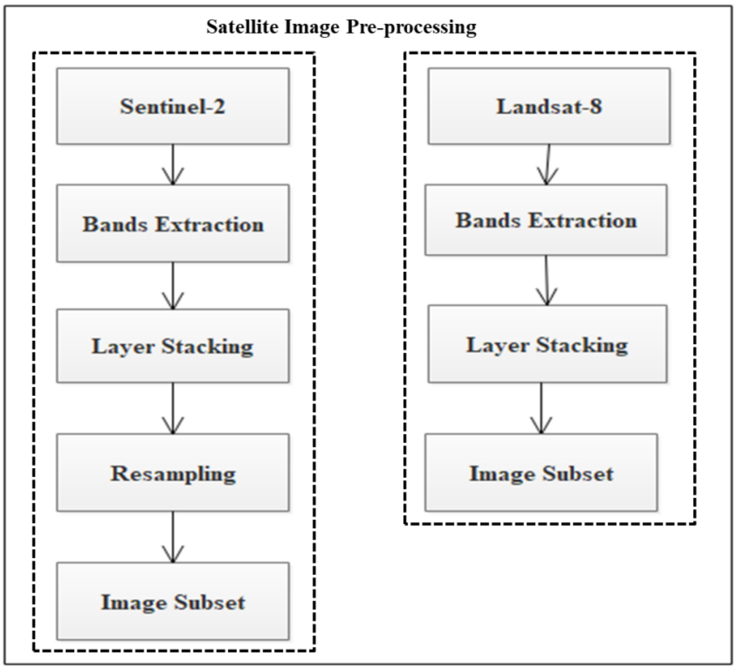


Figure 2 Preprocessing of Satellite Images

In Sentinel-2, the first step is to extract the NIR, Red, and Green bands from an image and then layered stacks of all the extracted bands in a single layer. The layer of bands was resampled using the Blue band of resolutions because the resolution of the Blue band is 10 meters. In Landsat-8, TIRS bands have been extracted from the images and then layer stacking is applied to take in a single layer. Subsets from both the images have been taken as per the undertaken study area.

Satellite Image Classification

Satellite images have become popular due to classification and change detection in metropolitan areas [10]. Numerous applications have made use of classification approaches, including agricultural monitoring, wetland mapping, and forest cover mapping discrepancies, land cover change detection, natural disaster assessment, and water resource applications [10]. Image classification is a significant method for managing and organizing a lot of image data that can address the issue of disordered image data [26]. Satellite image classification is an effective method for extracting information from a satellite image. It is the process of categorizing pixels into relevant classes/labels such as urban, barren land, water bodies, etc. [27]. Satellite image classification is a not difficult task; the analyst must have numerous options and choices during the process [8]. Many statistical learning techniques are currently used to drive image classification [9]. The satellite image classification is responsible for categorizing various objects into matching categories based on their features [9]. Image feature extraction is an important part of image classification which is different from computer vision and image operations [26]. Feature extraction techniques have been used in a variety of applications such as the reconstruction and preservation of the ancient building, satellite image analysis, urban planning, and medical diagnostics [26]. The most commonly used supervised classification techniques are maximum likelihood (ML), artificial neural network (ANN), and support vector machine (SVM), which have been used by many researchers for various applications.

In order to classify image pixels, the maximum likelihood (ML) classification method takes their probability value for classification; this technique develops a set of hyper-planes in a high-dimensional space [9]. The maximum likelihood technique is a statistical supervised technique for pattern recognition that uses the probability values of the image’s pixels to assign the relevant classes. For classifying satellite image pixels, maximum likelihood is an effective strategy. However, it takes time, and outcomes are often poor due to a lack of ground truth data [8]. Support Vector Machine (SVM) is used as a non-parametric technique with a binary classifier to handle more input data efficiently. The hyper-plane selection and kernel parameters determine the accuracy and performance of this technique. SVM's structure is more difficult to compare than other techniques and has a poor level of result trenchancy [22]. The SVM reduces the amount of required training data and increases the classifier's speed significantly. The main strength of SVM is its kernel illustration which allows for non-linear mapping of the input space to feature space. Some of the most popular SVM kernels are the Linear kernel, Polynomial kernel, and Gaussian kernel [22]. The performance of the SVM is determined by the kernel type and size chosen. The radial basis function kernel is used in many studies, and it has been proven to be more suitable for land use land cover applications [15]. An artificial neural network (ANN) classifier is a non-parametric technique for image classification whose performance and accuracy are determined by the network's structure and inputs [9]. The inputs are fast in this classification but the training process is slow, and architecture selection is also a difficult task in the neural network classification [22]. ANN computing models are inspired by biological neural networks that offer new ways to handle difficulties that arise during natural tasks [12]. An ANN classifier contains some human brain processes as well as the natural desire to store experimental information [9]. ANN has a collection of sequence layers, and each layer of the neural network has a large number of neurons. The weighted connection is tightly connected to all the neurons for processing of the following layers [9]. The weight of a link between two nodes is used to determine how much one node will influence the other [12]. In this study, various supervised classification techniques such as maximum likelihood (ML), artificial neural network (ANN), and support vector machine (SVM) have been used to classify urban areas effectively and efficiently.

Principal Component Analysis (PCA)

Principal component analysis (PCA) is a method for reducing complex data or dimensions in multispectral datasets. PCA is a statistical tool used to analyses signals in multivariate approaches such data reduction, optimal representation, and display of multispectral satellite images [21]. PCA is a statistical technique for reducing a set of associated multivariate variables in satellite remote sensing, and opened up new research avenues in mapping, inventorying, and monitoring the earth's natural resources in digital format [21]. The PCA has been used with satellite images for different purposes [28]. When classifying the land cover with satellite images, the PCA successfully observed the linear combination of data frames with human interpretability in various spatial, spectral, and temporal variations [21]. In the case of land study regions, PCA has been determined to reduce the dimensionality of the dataset in the categories of all forms of urban, soil, water, rocks, forest, and open fields, and it is a more powerful mathematical computing technique tool for feature extraction [21]. This technique is used for determining the correctness of a multispectral satellite image processing system. PCA can increase the accuracy of satellite images by enhancing classification efficiency in terms of high-speed processing time and data categorization. The interpretation of dimensionality reduction with image visualization is achieved using the mathematical matrix technique [21]. The PCA is based on the fact that adjacent bands of the hyperspectral image, and provides highly correlated and frequently nearly identical information about the item. The analysis is used to alter the original data to remove the bands' correlation [28].

Land Surface Temperature (LST)

The components of weather and climate that include factors like temperature, rainfall, air pressure, and humidity are crucial components of the Earth's system. The average weather is commonly used to define climate [29]. Land use land cover change mapping has been used to track environmental changes as a crucial component of studying the earth's surface land cover in the field of climate change events like floods and droughts [11]. Urban planning, land-cover mapping, and environmental monitoring are all made possible in low cost, fine-scale data using satellite images [11]. Land surface temperature (LST) is a crucial factor in numerous scientific studies due to its effect on the interactions between the atmosphere and the land [24]. In several domains, including climate change, urban land use/cover, and heat balance studies, LST is a significant input for climate models and a key factor [25]. LST is the Earth's surface temperature that comes into direct contact with the measuring device (usually measured in Kelvin) [25]. LST is the temperature at which heat and radiation from the Sun are absorbed, reflected, and refracted onto the surface of the Earth's crust. LST varies with changes in weather conditions and other human activity which makes it difficult to accurate predictions of land temperature [25]. LST is the most important factor in determining a location's highest and lowest temperature. Land cover and vegetation mapping at the regional local scale can be done with medium spatial resolution data from satellites such as Landsat and SPOT. The Thermal Infrared Sensor (TIRS) and Operational Land Imager (OLI) on board Landsat-8 can be used to gauge the temperature of the land surface [25]. Using data from Landsat missions (Landsat 5, 7, and 8), the Mono Window Algorithm (MWA), Radiative Transfer Equation (RTE) technique, Single Channel Algorithm (SCA), and Split Window Algorithm (SWA) were tested as LST retrieval methods over rural pixels [24]. In comparison to traditional approaches, the satellite image is a major source of Earth observation and it allows large-scale work with affordable, accurate (depending on the research design), and speedier findings [24].

The temperature of the Earth's surface is represented by LST which is one of the fundamental elements that affects surface energy balance, regional climates, heat fluxes, and energy exchanges [24]. Meteorological station radiance observations can be used to determine LST; it is a point-based measurement that does not often allow for large-scale monitoring. TIR band of satellite image enables large-scale monitoring even worldwide temporal and spatial LST investigations have been done by TIR band of satellite image [24]. There are three ways used to evaluate LST values from satellite images such as the Temperature-based method (T-based), the Radiance-based method (R-based), and cross-validation. Many researchers have presented alternative methodologies for LST retrieval considering these parameters [24]. For more than four decades, the Landsat series of satellites has consistently produced space-based moderate-resolution satellite images. The Landsat has launched its eighth generation of satellite and the mission of Landsat's earth observation started on July 23, 1972 [24]. According to the overlay of the transect profiles drawn on the land use/land cover change map over the land surface temperature map, the land surface temperature has increased in those areas that have undergone changes from open forest to paddy, open forest to settlement, paddy to settlement, and deposit to settlement [30]. Landsat with their medium-resolution is better suited to monitoring surface temperature in cities or smaller areas [31]. Change detection using satellite images has traditionally relied on post-classification or multitemporal classifications [32]. In this study, the land surface temperature measured by TIR bands of Landsat-8 image and the formula of LST measurement is mentioned in Eqn. 1.

$$\mathbf{LST} = \frac{\left(\left(\dfrac{\mathbf{k2B10}}{\mathbf{alog}\left(\dfrac{\mathbf{k1B10}}{\mathbf{0}}.\mathbf{00033442} * \mathbf{b10} + \mathbf{0.1}\right)} + \mathbf{1}\right) + \dfrac{\mathbf{k2B11}}{\mathbf{alog}\left(\dfrac{\mathbf{k1B11}}{\mathbf{0.00033442} * \mathbf{b11} + \mathbf{0.1}} + \mathbf{1}\right)}\right) - \mathbf{273.15}}{\mathbf{2}} \quad (\mathbf{1})$$

Where, K2B10 and K1B10 are a constant of band 10, and K2B11, and K1B11 are the constant of band 11, the values of constant mentioned on the metafile of the image. B10 and B11 are the cropped image of band10 and band

Regression and Correlation

Regression and Correlation are two separate strategies that are not mutually incompatible. In general, regression is used for prediction, and correlation is used to establish the degree of link [33]. Although Regression and Correlation are closely related and both are distinct concepts.

Regression

Regression analysis is a mathematical technique that examines the relationship between dependent and independent variables [34]. Independent factors are referred to as repressor variables or explanatory variables, while dependent variables are referred to as target variables regressed variables, or research variables [34]. For detecting and estimating trends in environmental variables of interest, a variety of statistical techniques are available such as simple correlation and regression studies, time-series analysis, non-parametric statistics methods, etc [35]. The most important prerequisite for a successful firm is future forecasting opportunities and risk estimation. Regression analysis may be used for a lot more than forecasting. A statistical technique that looks at the linear relationship between two or more quantitative variables is the linear regression analysis approach [34]. Using historical data, linear regression analysis seeks to predict future outcomes and identify linear relationships [34].

Correlation

The correlation is a concept that has far more applications than merely 2D scatterplots. There is also "multiple correlations," which is when numerous independent variables are correlated with a single dependent variable. There's also "partial correlation," which is the relationship between two variables while controlling for third or more variables [33]. The coefficient of correlation is a mathematical formula used to measure relationships between two variables. The intensity and direction of the association between two variables are described by correlation. It emphasized that the correlation coefficient between two variables only measures a linear relationship [34]. There are a variety of correlation coefficients to handle the unique qualities of dichotomies and other sorts of nominal and ordinal variables, as well as other measures of relationship (and for time-series analysis as well) [33]. In this study, the Pearson correlation coefficient has been calculated, and the formula is mentioned in Eqn. (2).

$$\mathbf{correlation}\ (\mathbf{r}) = \frac{\mathbf{N}\sum\mathbf{XY} - (\sum\mathbf{X})(\sum\mathbf{Y})}{\sqrt{[\mathbf{N}\sum\mathbf{X^2} - (\sum\mathbf{X})^2][\mathbf{N}\sum\mathbf{Y^2} - (\sum\mathbf{Y})^2]}} \qquad (\mathbf{2})$$

Where, N is the number of pairs of scores, Σxy is the sum of the products of paired scores, Σx is the sum of x scores, Σy is the sum of y scores, Σx2 is the sum of squared x scores, Σy2 is the sum of squared y scores.

Model Development

Human actions have a huge impact on the terrestrial surface of the earth. The Earth's surface is used to fulfil the demands of humans such as food production, energy extraction, urban expansion, etc. Human activities are reducing the Earth's vegetation cover continually, and carbon dioxide concentrations in the atmosphere are rising. The surface energy budget illustrates its effect and as a result local, regional, and global climates change. The objective of this study is to estimate the Earth's surface changes, and also measure the Earth's surface temperature. In this study, a semi-supervised deep learning model has been developed that is capable to monitor the Earth's surface such as urban areas, barren lands, water bodies, and agricultural fields, and also measuring the land surface temperature. The model provides the correlation between urban areas and land surface temperature as a result. The detailed diagram of the model is shown in Figure 3.

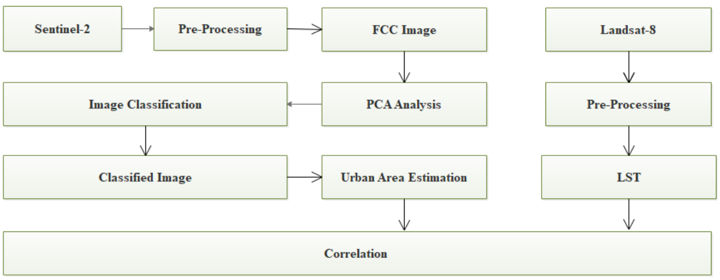
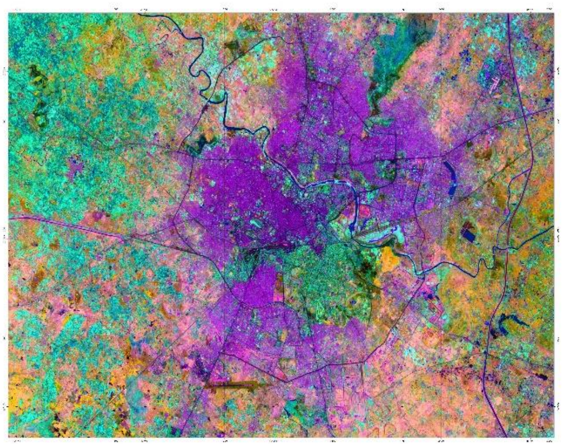


Figure 3 Semi-Supervised Deep Learning Model for Finding Correlation between Urban Areas and Land Surface Temperature

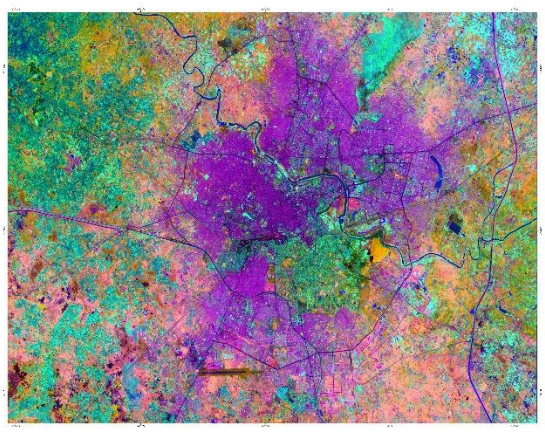
Copernicus Open Access Hub are providing Sentinel-2 images for earth surface monitoring, and these images must be preprocessed before being analyzed and interpreted for earth surface monitoring. The NIR, Red, and Green bands of the image were used to create the FCC image after preprocessing. The PCA has been applied to each FCC image for dimension reduction and then applied different supervised image classification techniques on the PCA image. The obtained classified images help to estimate urban areas, while Landsat-8 images have been downloaded from Earth Explorer for land surface temperature (LST) measurement. After downloading the images, the downloaded image needs to be preprocessed. To estimate the LST, the LST formula was applied to a Landsat-8 image that had already been processed. The acquired urban area and LST are used to examine their correlation.

Result and Discussion

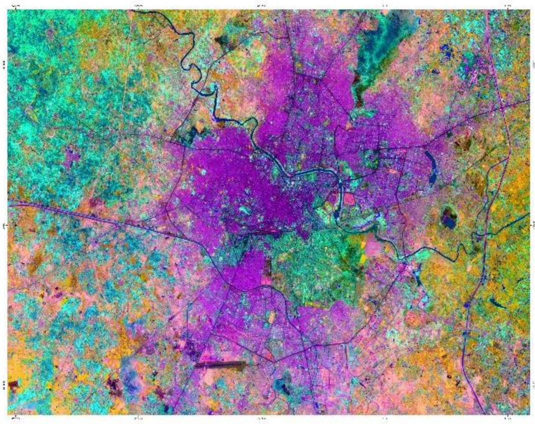
The developed model can easily estimate the urban areas and land surface temperature, the obtained urban area and land surface temperature help to find the correlation between urban areas and temperature that provide the model as a result. In this paper, Sentinel-2 images are used for urban area estimation and Landsat-8 images are used for land surface temperature measurement and the details of satellite images are given in Table 1 and 2. Sentinel-2 image has multiple bands, each band has a different resolution hence two or more bands with different resolution cannot used together that making it difficult to interpret and analyses of the Earth's surface. To resolve this issue, image resampling is used to converts each band's resolution into the same format that best suits the study's requirements. Hence, false color composite (FCC) images of Img1 to Img5 images have been generated by assigning the red color to the NIR band, green color to the Red band, and blue color to the Green band images and to reduce the dimension reduction of the data, PCA has been applied on each FCC image. The retrieved PCA images from Img1 to Img5 are shown in Figure 4 (a) to 4 (e).



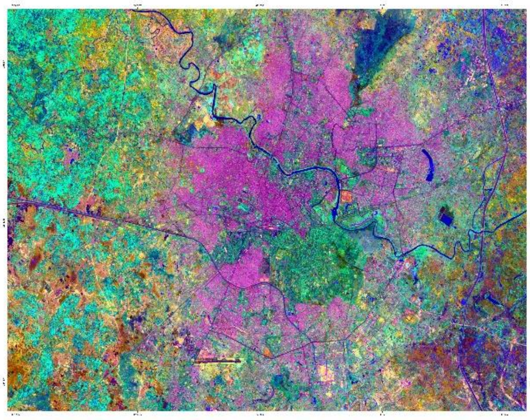
(a) PCA Image of Img1



(b) PCA Image of Img2



(c) PCA Image of Img3



(d) PCA Image of Img4

(e) PCA Image of Img5

Figure 4. PCA Images of Study Area

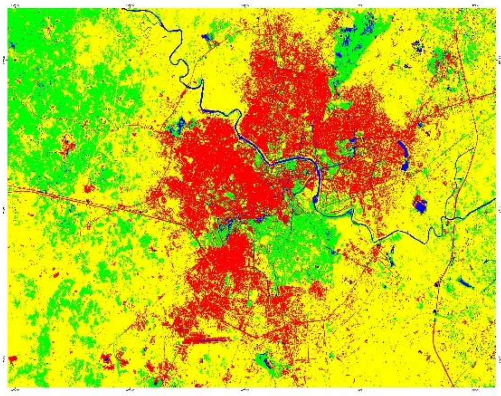
Images retrieved by PCA are used for land cover classification by using the various classification techniques such as maximum likelihood (ML), artificial neural network (ANN), and support vector machine (SVM) for finding a suitable classifier for this study that can produce efficient classified image. The obtained classification accuracy (overall accuracy and kappa coefficient) through ML, ANN, and SVM are shown in Table 3.

Table 3. Classification accuracy retrieved through ML, ANN and SVM.

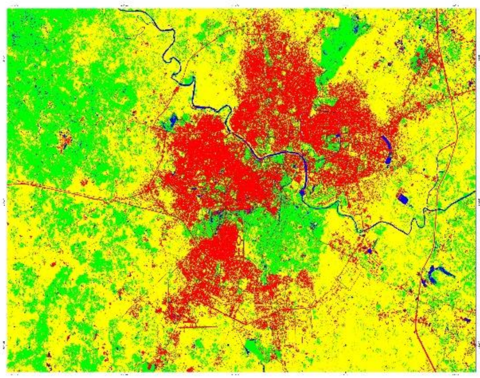
Image ID	Algorithm	Kappa	Overall Accuracy
Img1	ML	0.90	92.70%
	ANN	0.91	93.45%
	SVM	0.92	94.01%
Img2	ML	0.91	93.26%
	ANN	0.91	93.26%
	SVM	0.91	93.4%5
Img3	ML	0.83	87.08%
	ANN	0.88	91.20%
	SVM	0.88	91.39%
Img4	ML	0.79	84.08%
	ANN	0.80	85.02%

	SVM	0.80	85.21%
Img5	ML	0.89	91.57%
	ANN	0.88	91.01%
	SVM	0.88	91.76%

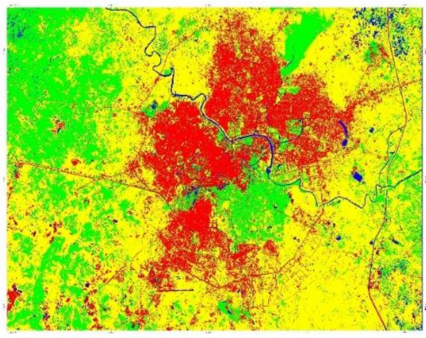
In this study, it is observed that the result of SVM is more accurate to the ML and ANN Classifier because the kappa and overall accuracy of SVM are greater than others. The SVM classified images of Img1 to Img5 are shown in Figure 5 (a) to (5e), where urban is represented by red color, agriculture fields are represented by green color, water is represented by blue color, and barren land is represented by yellow color.



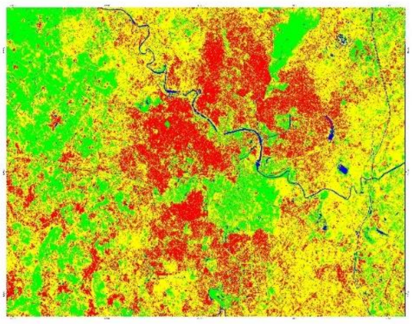
(a) Classified Image of Img1



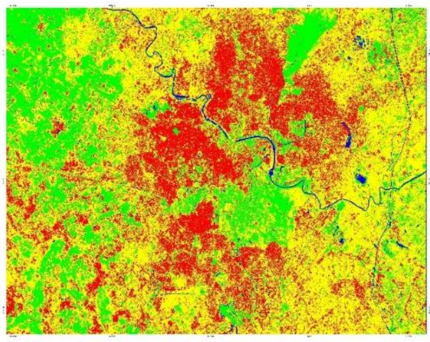
(b) Classified Image of Img2



(c) Classified Image of Img3



(d) Classified Image of Img4



(e) Classified Image of Img5

Figure 5 SVM Classified Images of Study Area

Class-wise classification accuracy retrieved through SVM classifier is shown in Table 4.

Table 4. Class-wise classification accuracy.

Image ID	Urban (%)	Agriculture (%)	Water (%)	Barren (%)
Img1	18.961%	18.339%	1.930%	60.770%
Img2	19.138%	23.427%	1.720%	55.716%
Img3	19.389%	22.257%	2.494%	55.861%
Img4	23.972%	20.785%	1.731%	53.512%
Img5	24.434%	21.706%	2.573%	51.287%

From Table 4, it is observed that there is a slight increase in urban areas every year. Further, the actual urban area has been estimated as shown in Table 5, and the bar chart is shown in Figure 6.

Table 5. Urban Area Estimation

Image ID	Urban Area (Hectare^2)
Img1	15626.53
Img2	15771.91
Img3	15978.61
Img4	19755.64
Img5	20136.93

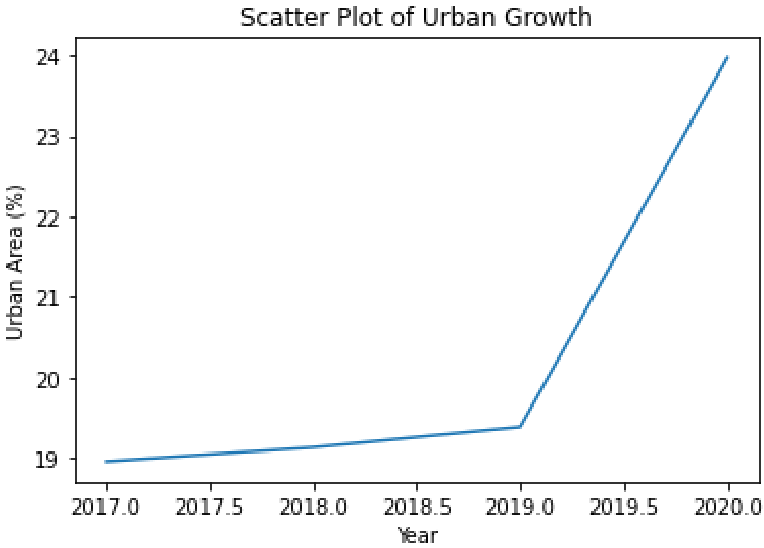


Figure 6. Urban Area **Estimation (2017-2021).**

After the urban area estimation, another important part of this study is the land surface temperature (LST) measurement. Landsat-8 images have been used for LST measurements, and the details of the used images are mentioned in Table 1. After the preprocessing of Landsat-8 image, LST has been measured by using the Eqn. (1) and the retrieved LSTs from img6 to img10 are shown in Table 6.

Table 6. Land Surface Temperature.

Image ID	LSTmin	LSTmax	LSTmean
Img6	38.35	43.20	41.07
Img7	39.52	42.73	41.33
Img8	24.25	47.34	41.76
Img9	27.12	46.94	42.89
Img10	41.31	44.92	43.06

After the urban area estimation and land surface temperature measurement, there is a need to prepare a dataset of the values of the estimated urban area and land surface temperature for finding the correlation between them. The formula of the Pearson correlation coefficient is mentioned in Eqn. (2), and the prepared dataset is mentioned in Table 7.

Table 7. Correlation between Urban Area and Land Surface Temperature

Year	Urban Area (X)	LST (Y)	XY	X^2	Y^2
2017	18.96	41.06	778.49	359.48	1685.92
2018	19.13	41.32	790.45	365.95	1707.34
2019	19.38	41.75	809.11	375.58	1743.06
2020	23.97	42.88	1027.83	574.56	1838.69
2021	24.43	43.06	1051.95	596.82	1854.16
	∑X =105.87	∑Y = 210.07	∑XY = 4457.83	∑X2 = 2272.39	∑Y2 =8829.17

From Table 7, values of ∑X, ∑Y, ∑XY, ∑X2, and ∑Y2 are used in Eqn. 2 for finding the correlation.

$$\boldsymbol{correlation}\ (\boldsymbol{r}) = \frac{\mathbf{5}(4457.83) - (105.87)(210.07)}{\sqrt{[\mathbf{5} * 2272.39 - 105.87^{\mathbf{2}}][\mathbf{5} * 8829.17\ -\ 210.07^{\mathbf{2}}]}}$$

$$\boldsymbol{r} = \frac{49.03}{\sqrt{2523.37}}$$

$$r = 0.97$$

The obtained value of r is 0.97 hence it proves that there is a positive correlation between the urban area and land surface temperature which defines the urban areas increasing concurrently with temperature also increasing. The correlation is shown in Figure 7, and the bar chart is shown in Figure 8.

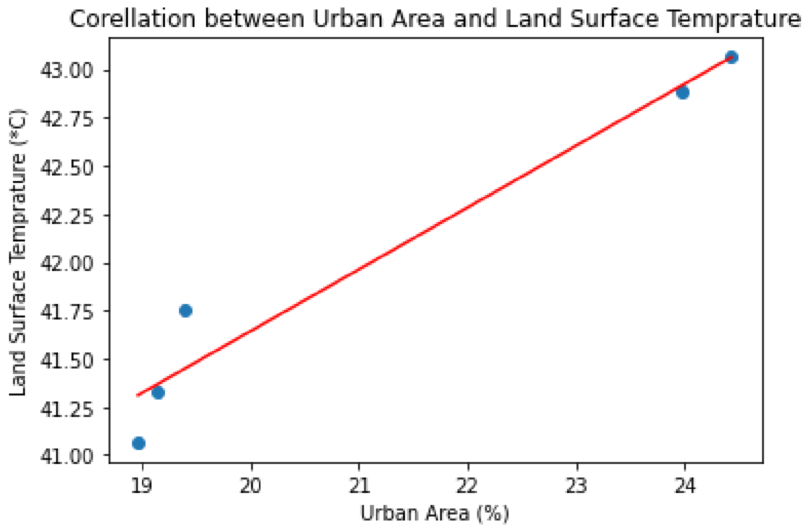


Figure 7. Correlation of Urban Area and Land Surface Temperature

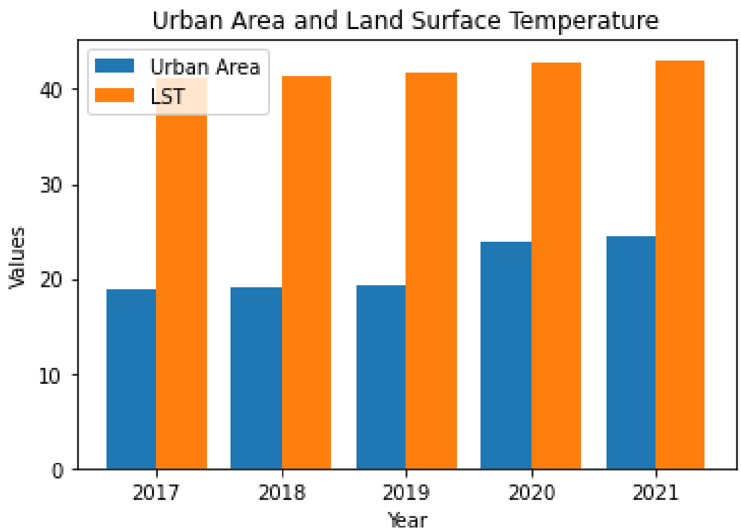


Figure 8 Bar Chart of Urban Area and Land Surface Temperature

Trend Analysis

Trend analysis is used to predict the future movements of a particular variable using the statistical approach. In other words, it is a method that aims to predict future behaviors by examining past data. The obtained result (urban area and land surface temperature) of this study is used as a dataset that helps to predict the trend of urban and land surface temperature. In this analysis, predict the urban area and temperature for the 10 years (2022 to 2031) using the linear trend analysis technique. The obtained result (urban area and land surface temperature) are mentioned in Table 6. The formula of Linear Trend Analysis is mentioned in Eqn. (3).

$$\mathbf{Y=MX+C} \quad \mathbf{(3)}$$

Where Y is the dependent variable, X is the independent variable, M is the slope, and C is the intercept.

The analysis is divided into two-part, the first part predicts the urban areas, and the second part predicts the LST (mean) for the 10 years. The urban area prediction formula is mentioned in Eqn 4.

$$\textbf{Urban area=M*Year +C} \quad \mathbf{(4)}$$

The value of M, C, and accuracy are mentioned in table 8

Table 8. Slope, Intercept, Accuracy of the first part.

Slope (M)	Intercept (C)	Accuracy
1.578	-3164.803	81.14

The urban growth area of the 10 years from 2022 to 2031 that is predicted by linear trend analysis is shown in Figure 10.

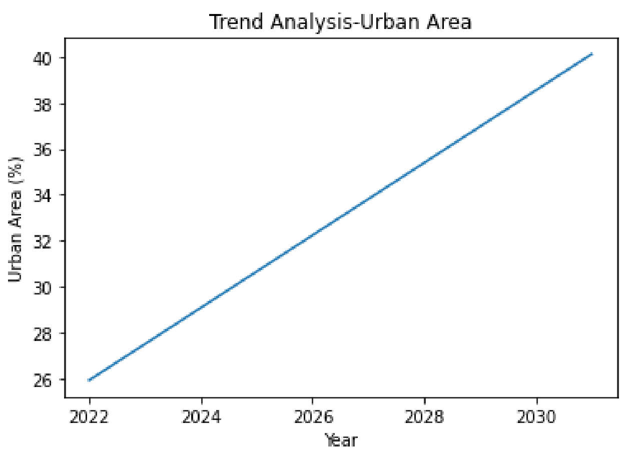


Figure 10. Urban Area Prediction

In the second part, applied the linear trend analysis on obtained result for finding temperature growth based on urban growth hence the dependent variable will be temperature and the independent variable will be an urban area. The formula of Temperature Prediction is mentioned in Eqn. (4)

$$\textbf{Temperature=M*Urban Area +C} \quad \textbf{(4)}$$

The accuracy, slope, and intercept is mentioned in table 9.

Table 9. Slope, Intercept, Accuracy of Second Part

Slope (M)	Intercept (C)	Accuracy
0.319	35.243	91.90

After the training of temperature prediction, the obtained urban areas from the urban area analysis help to predict the temperature in this analysis, and the predicted temperature with urban areas is mentioned in Figure 11.

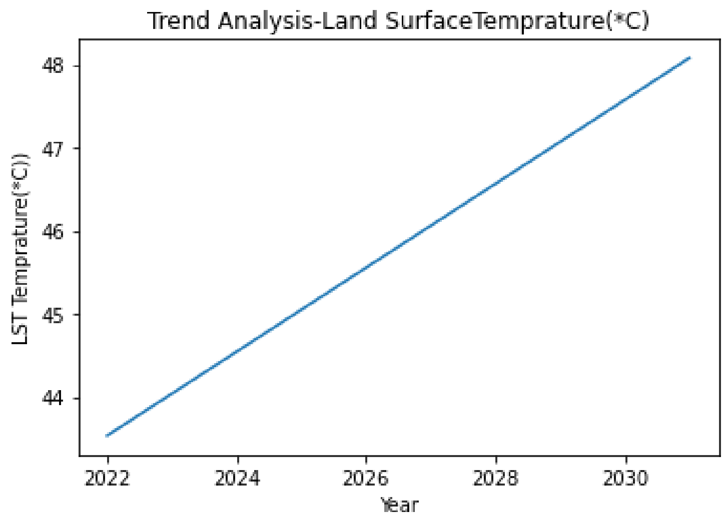


Figure 11. Temperature Prediction

After the urban area prediction and temperature prediction, both are plotted together and the combined graph is shown in Figure 12.

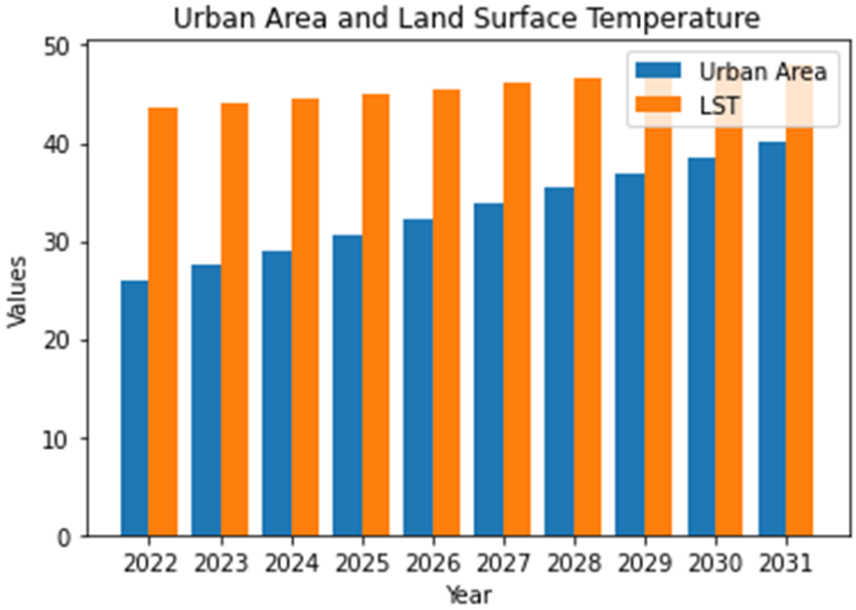


Figure 12. Urban area and LST Prediction for the upcoming 10 **years (2022 to 2031).**

From Fig 10, it is observed that the urban areas increasing per year, and land surface temperature is also increasing as per the urban area. After 10 years, the urban areas of the study area will reach approx. 40% and the temperature will be approx. 48*C.

Conclusion

The satellite image is a powerful source of earth observation, lots of frameworks, methods, and techniques have been developed that are providing lots of meaningful information about the earth's surface. The information extraction from satellite image is not an easy task, satellite image holds multiple bands with different resolutions. The earth's surface is specifically described in each band of the image. Earth surface monitoring has made extensive use of machine learning methods and satellite pictures. Satellite image has heterogeneous features so the machine learning algorithms are not so much reliable with satellite images. In this study, a semi-supervised deep learning model has been developed to find out the correlation between the urban area and land surface temperature, in which Sentinel-2 images have been used for urban area estimation and the Landsat-8 images have been used to estimate land surface temperature. The estimated values of this study have been used to calculate the correlation between them, and also used to predict the trend of the urban area and land surface temperature for the 10 years (2022 to 2031) using the linear trend analysis technique, and it is observed that the LST has been increasing in the taken period, according to trend analyses. In the future, the developed model will be very helpful to earth surface monitoring and land surface temperature measurement for which some corrective measures may be taken. A further way to lower the temperature is to disperse metropolitan residents over increasing built-up areas toward the periphery. The consequences of increasing temperatures can be lessened by maintaining open space and removing more tangible structures. It follows that there is a dire requirement on behalf of ongoing monitoring of city-level dynamics of LU/LC, as well as for the development of viable and systematic built-up terrestrial usage regulations, in order to keep tabs on the phenomena of urban heat island (UHI) intensification.

Acknowledgement

The authors are very thankful to the Advanced Computing Research Laboratory, Department of Computer Applications, Integral University Lucknow, India because it provided the necessary support to carry out this work. The MCN No. provided by the University is IU/R&D/2022-MCN0001614 for this work.

References

[1] X. Zhao , X. Li , Y. Zhou , and D. Li, "Analyzing Urban Spatial Connectivity Using Night Light Observations: A Case Study of Three Representative Urban Agglomerations in China," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1097–1108, 2020.

[2] M. Sahana, H. Hong, H. Sajjad, "Analyzing urban spatial patterns and trend of urban growth using urban sprawl matrix: A study on Kolkata urban agglomeration, India," *Ecol. Sociobiol*, vol. 628, pp. 1557–1566, 2018.

[3] J. Chadchan, R. Shankar, "An analysis of urban growth trends in the post-economic reforms period in India," *International Journal of Sustainable Built Environment*, vol. 1(1), pp. 36–49, 2012.

[4] K. Li, Y. Chen, "A Genetic Algorithm-Based Urban Cluster Automatic Threshold Method by Combining VIIRS DNB, NDVI, and NDBI to Monitor Urbanization," *Remote Sensing*, vol. 10(277), pp. 1–21, 2018.

[5] H. Li, Q. Li, G. Wu, J. Chen, and S. Liang, "The Impacts of Building Orientation on Polarimetric Orientation Angle Estimation and Model-Based Decomposition for Multilook Polarimetric SAR Data in Urban Areas," *IEEE Transactions on Geoscience And Remote Sensing*, vol. 54(9), pp. 5520–5532, Sept. 2016.

[6] Z. Mitraka, N. Chrysoulakis, G. Doxani, F. Frate, M. Berger, "Urban Surface Temperature Time Series Estimation at the Local Scale by Spatial-Spectral Unmixing of Satellite Observations," *Remote Sensing*, vol. 7(4), pp. 4139–4156, 2015.

[7] T. Ahmed, D. Singh, "Probability density functions based classification of MODIS NDVI time series data and monitoring of vegetation growth cycle," *Advances in Space Research* , vol 66(4), pp. 873-886, 2020.

[8] S. Abburu, S. Golla, "Satellite Image Classification Methods and Techniques: A Review," *International Journal of Computer Applications*, vol. 119(8), pp. 20–25, January 2015.

[9] S. Srivastava, T. Ahmed, "Feature-Based Image Retrieval (FBIR) System for Satellite Image Quality Assessment Using Big Data Analytical Technique," *Psychology And Education*, vol. 58(2), pp. 10202–10220, 2021.

[10] A. Rai, N. Mandal, A. Singh, K. Singh, "Landsat 8 OLI Satellite Image Classification using Convolutional Neural Network," Procedia Computer Science, presented at the 124th Annu. Meeting American Public Health Association, New York, USA, 1969.

[11] A. Jamali, “Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: a case study of the northern region of Iran,” *International Journal of Computer Applications, SN Applied Sciences*, doi: https://doi.org/10.1007/s42452-019-1527-8.
[12] S. Srivastava, T. Ahmed, “Similarity-Based Neural Network Model for FBIR System of Optical Satellite Image Quality Assessment,” *Turkish Journal of Computer and Mathematics Education*, vol. 12(13), pp. 2391–2410, 2021.
[13] C. Gómez, J. White, M. Wulder, “Optical remotely sensed time series data for land cover classification: A review,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 116, pp. 55–72, 2016.
[14] S. Bhatti, N. Tripathi, “Built-up area extraction using Landsat 8 OLI imagery,” *GIScience & Remotef Sensing*, vol. 51(4), pp. 445–467, July 2014.
[15] M. Cavur, H. Duzgun, S. Kemec, D. Demirkan, “Land Use And Land Cover Classification Of Sentinel 2-A: St Petersburg Case Study,” *St Petersburg Case Study, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 13–16, 2019.
[16] M. Steinhausen, P. Wagner, B. Narasimhan, B. Waske, “Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions,” *Int J Appl Earth Obs Geoinformation*, vol. 73, pp. 595–604, 2018.
[17] T. Liu, X. Yang, “, Monitoring land changes in an urban area using satellite imagery,” *Applied Geography*, vol. 56, pp. 42–54, 2015.
[18] G. Yiqiang, W. Yanbin, J. Zhengshan, W. Jun, Z. Luyan, “Remote sensing image classification by the Chaos Genetic Algorithm in monitoring land use changes,” *Mathematical and Computer Modelling*, vol. 51, pp. 1408–1416, 2010.
[19] S. Jore, P. Badadapure, “Remote Sensing Image Segmentation using Linear Regression,” *International Journal of Engineering Research & Technology*, vol. 3(4), pp. 2279–2282, 2014.
[20] H. C. Verma, T. Ahmed, S. Rajan, “Mapping and Area Estimation of Mango Orchards of Lucknow Region by Applying Knowledge Based Decision Tree to Landsat 8 OLI Satellite Images,” *International Journal of Innovative Technology and Exploring Engineering* , vol 9(3), pp. 3637-3635, 2020.
[21] S. Hiremath, D. Kulkarni, “Advances in Principal Component Analysis in Multispectral Satellite Imagery,” *international Journal of Engineering Sciences and Management- A Multidisciplinary Publication of VTU*, vol. 1(1), pp. 39–48, 2019.
[22] S. Srivastava, “A survey on satellite image classification approaches,” *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 6(2), pp. 480–483, 2020.
[23] D. Zakari, B. Joda, K. Yahya, “Satellite Image Enhancement Using Principal Component Analysis (PCA) Transformation Technique to Maximize the Signal-to-Noise ratio for Hyper-Spectral Data,” *International Journal of Application or Innovation in Engineering & Management*, vol. 10(7), pp. 51–61, 2021.
[24] A. Sekertekin, S. Bonafoni, “Land Surface Temperature Retrieval from Landsat 5, 7, and 8 over Rural Areas: Assessment of Different Retrieval Algorithms and Emissivity

Models and Toolbox Implementation," *remote sensing*, *doi:*https://doi.org/10.3390/rs12020294.

[25] Anandababu D, Purushothaman B M, D. Babu, "Estimation of Land Surface Temperature using LANDSAT 8 Data," *International Journal of Advance Research*, vol. 4(2), pp. 177–186, 2018.

[26] Y. Sun, L. Li, L. Zheng, J. Hu, W. Li, Y. Jiang, C. Yan, "image classification base on PCA of multi-view deep representation," *J. Vis. Commun. Image R*, vol. 62, pp. 253–258, 2019.

[27] H. C. Verma, S. Rajana , T. Ahmed, "A Review on Land Cover Classification Techniques for Major Fruit Crops in India- Present Scenario and Future Aspects," *In conference: International Conference on Sustainable Computing in Science, Technology & Management., Amity University Rajasthan, Jaipur-India, pp. 1591-1596, 2019.*

[28] C. Rodarmel, J. Shan, "Principal Component Analysis for Hyperspectral Image Classification," *Surveying and Land Information Science,* vol. 62(2) , pp. 115–122, 2002.

[29] A. Panda, N. Sahu, "A survey on satellite image classification approaches," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 6(2), pp. 480–483, 2020.

[30] M. Sahana, R. Ahmed, H. Sajjad, "Analyzing land surface temperature distribution in response to land use/land cover change using split window algorithm and spectral radiance model in Sundarban Biosphere Reserve, India," *Modeling Earth Systems and Environment*, doi:10.3390/rs5115969.

[31] G. Kaplan, U. Avdan, Z. Avdan, "Urban Heat Island Analysis Using the Landsat 8 Satellite Data: A Case Study in Skopje, Macedonia," *Proceedings*, vol. 2(7) , pp. 1–5, 2018.

[32] L. Martínez, V. Palà, R. Arbiol, L. Pineda, M. Joaniquet, ", Urban Change Detection on Satellite Images Series," *Application to Catalunya Area, doi: 10.1109/URS.2007.371834.*

[33] A. Asuero, A. Sayago, and A. Gonz´alez, "The Correlation Coefficient: An Overview," *Critical Reviews in Analytical Chemistry*, vol. 36, pp. 41–59, 2006.

[34] A. Pant, R. Rajput, ", Linear Regression Analysis Using R for Research and Development," *In book: Writing Qualitative Research Paper of International Standard., pp.180-195, Oct. 2019.*

[35] A. Hess, H. Iyer, W. Malm, "Linear trend analysis: a comparison of methods," *Atmospheric Environment*, vol. 35(30), pp. 5211–5222, October 2001.

[36] T. Miyauchi, T. Setoguchi, T. Ito, "Quantitative estimation method for urban areas to develop compact cities in view of unprecedented population decline", Cities, vol. 114, pp. 1-11, May 2022.

[37] T. Somokanta, "Urban climate change experiments in Gandhinagar, India", Urban Climate, vol. 43, May 2022.