CLASSIFICATION OF THE STRUCTURE OF CITIES THROUGH MID-RESOLUTION SATELLITE IMAGERY AND PATCH BASED NEURAL NETWORKS

Deepank Verma [1, *], Arnab Jana [1], Krithi Ramamritham [2]

[1] Centre for Urban Science and Engineering, Indian Institute of Technology Bombay - (deepank,arnab.jana)@iitb.ac.in
[2] Dept. of Computer Science and Engineering, Indian Institute of Technology Bombay – krithi@cse.iitb.ac.in

Commission V, SS: Emerging Trends in Remote Sensing

KEY WORDS: Convolutional Neural Networks (CNN), Autoencoders, Sentinel 2B, Indian Cities, t-SNE, Unsupervised Clustering.

ABSTRACT:

The studies in the classification of the urban spatial structure have been essential in deriving insights into the land cover and the built typology which helped in the estimation of energy consumption patterns, urban density, compactness, and hierarchy of settlements. However, the analysis and comparison of the physical forms of the cities have been attempted in a piecemeal fashion where the requirement of datasets and the computation power for analysis has been a major hindrance. With the advancement in machine learning based techniques, large datasets such as satellite imagery can be studied with advanced computer vision methods. These solutions may help in studying the intricate nature of human habitats in large extents of geographical areas including various urban areas. This study utilizes smaller patches of medium resolution Sentinel-2B Imagery of ten different cities in India to explore the urban forms present in these cities. This study uses Stacked Convolutional Autoencoder (CAE) to reduce the dimensionality of satellite imagery patches and unsupervised clustering techniques such as t-SNE and K-means to study the characteristics of similar patches. On analyzing the clusters through visual exploration, similar patches are delineated and provided with corresponding labels representing urban forms. Individual clusters are then studied with respect to each city. The motive of the study is to gain insights into the different types of morphological patterns present within and among cities.

1. INTRODUCTION

Land use Land cover (LULC) maps have been extensively used for the delineation of land characteristics. LULC maps are fundamental in the estimation of agricultural production (Zheng et al., 2015), analyzing biodiversity (Szostak et al., 2018), assessment of natural hazards (de Moel and Aerts, 2011; Khatami and Mountrakis, 2012) and urbanization (Taubenböck et al., 2009). The availability of open data from Earth Observation (EO) satellites combined with open image processing software and toolboxes have pushed the LULC based research to a new level. Several Machine learning algorithms have been studied and implemented to classify different features of the land (Noi and Kappas, 2018; Shao and Lunetta, 2012). However, the use of such algorithms in studying intra-urban features has been scarce. The major reason can be attributed to the availability of relatively coarser resolution of satellite imagery as open datasets which acts as a major limiting factor. Studies (Kuffer et al., 2017; Mboga et al., 2017) utilizing High and very high-resolution imagery have been conducted to understand the structure of cities which have limited availability with the researchers. The medium resolution datasets such as Sentinel and Landsat products, on the other hand, have only been utilized for regional level analysis.

With the rapid urban expansion due to urbanization particularly in cities in developing countries, there is a growing need to monitor changes in intra-urban structures and textures in quick succession. Urban morphology has been widely discussed in urban planning and management which studies form and function of urban areas. Morphology defines the uniqueness, identity and vibrant character of the city. Local Climate Zones (LCZ) (Bechtel et al., 2015) classification method has been widely used by researchers to understand the morphology and urban fabric. The methodology (Ching et al., 2015) to create LCZ maps involves manual delineation of training samples for supervised classification which specifically requires a field expert to interpret scenes and to create training samples. This study utilizes a novel unsupervised classification technique to categorize urban structure in which categorization of classes is done with the help of dimensionality reduction and clustering methods.

Research in image processing is undergoing a paradigm shift with the inclusion of state of the art Deep learning methods. Convolutional Neural Networks (CNN) are one of such methods which have provided near human accuracy in Computer Vision tasks such as image classification. CNNs can learn the hierarchical representation of the variety of features present in the spatial and spectral domain in satellite imagery. In this study, CNN is used as encoder and decoder to learn the feature representation and to obtain the reduced dimension of input data as embeddings.

This study aims to study the inherent structure of cities through unsupervised learning. It provides a step by step approach from the creation of patch-based dataset to the classification of results using CAE as a dimensionality reduction technique. The present methodology is created as a test, which may be scaled to include various cities or cities from different countries.

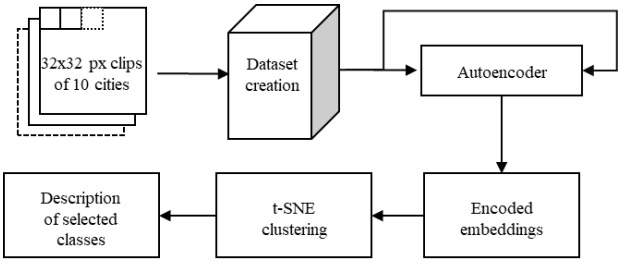

Figure 1: Flow diagram showing the methodology of the study.

The methodology includes the sequential clipping of image tiles from cities, which are further stacked together to form a training

* Corresponding author

dataset. The dataset is then passed through the CAE, and the embeddings for each of the clips are recorded and stored. These embeddings are plotted with the help of a t-SNE algorithm (Van Der Maaten and Hinton, 2008). Clusters are created through K-means algorithm. Each cluster is assigned a class after studying the gist of information it provides. The presence of such classes is determined in the cities and statistics is generated (Fig. 1).

2. DATA COLLECTION

Satellite imagery is downloaded for the ten largest cities in India from Sentinel 2B database (Table 1). All the imageries were captured within one month (from 15 Mar to 15 Apr 2018) of duration. Bands 2,3,4 and 8; which represents Blue, Green, Red, and Infra-Red regions of the spectrum are stacked to create a composite set for each of the cities. The City boundaries are used to clip the city extents from the downloaded image dataset. The large variation in areas of the considered cities is evident in Table 1. The city of Delhi has the largest area among the other cities, which is due to the inclusion of contiguous hinterlands which extends up to the border of State of Delhi. To create patches from each set of imagery we used GDAL to sub-divide each city into smaller images of size 32x32 pixels, each covering an area of 10.2 Ha. The process resulted in the generation of approximately 105 patches (Table. 1). The set of patches are further processed with the Convolution based Autoencoders and Unsupervised clustering algorithms. Vanilla clustering techniques such as Random forests and KNN are effective in identifying clusters in small sets of data. However, these algorithms are not feasible for the larger datasets (105x32x32x4 values). Commonly used dimensionality reduction algorithms such as Principal Component Analysis (PCA) perform better in unstructured datasets. However, for the structured datasets, such as satellite imageries we utilized CNN to extract relevant features and to reduce dimensionality from image tiles.

Table 1: Data preparation from 10 cities.

S.no	Cities	Area (in sq.km)	No. of 32x32 tiles
1	Ahmedabad	673.792	6580
2	Bengaluru	1342.2592	13108
3	Chennai	863.8464	8436
4	Delhi	3297.28	32200
5	Hyderabad	1501.0816	14659
6	Jaipur	725.9136	7089
7	Kolkata	1514.496	14790
8	Mumbai	1000.2432	9768
9	Pune	538.0096	5254
10	Surat	508.3136	4964
	Total	11962.6	116823

3. ARCHITECTURE OF CONVOLUTIONAL AUTOENCODER

CNN can be dubbed as the extension of regular Artificial Neural Networks which focuses on image analysis. CNN is primarily used for Image classification, segmentation and object detection tasks widely used in industry and academia. The property of CNN to learn features from the provided image array while preserving local structure and composition of the image is utilized considerably in this study. These networks discover features and patterns and frame meaningful representations from the input data with the help of various filters present in the hidden layers. We utilize the learned features with the help of Autoencoders to create a unique set of values (embeddings) for each patch.

Autoencoders perform automated translation of datasets from higher to a lower dimension known as embeddings. The autoencoders are data compression algorithm which is based on the three functions: encoding, decoding and the loss values between compressed and decompressed representation. The encoding task reduces the dimensionality of the input to generate its compressed representation. These compressed values are learned by the decoder to produce the output similar to the provided input. The Convolution Autoencoders (CAE) consist of convolutional layers as a part of encoders and decoders. CAE is used in this study to process the patches and to generate the set of values in lower dimensions.

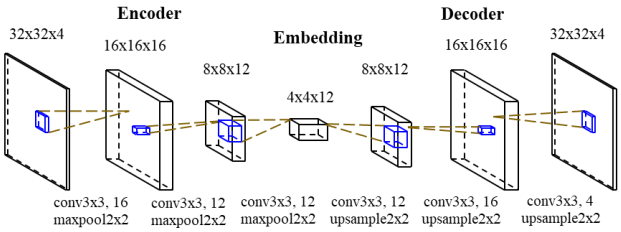

Figure 2: Architecture of Convolutional Autoencoder.

The CAE model (Fig. 2) includes convolution network (encoder), embedding layer and deconvolution network (decoder) which further consists of input, convolutional, max pooling, embedding, upsampling, and output layers in an organized fashion. The input layer is a placeholder for the prepared dataset in the form of 32x32x4 tiles. It is connected with a convolutional layer, in which kernel of size 3x3 is applied to the input layer, 16 activation maps are created and stacked to create a volume of 32x32x16, max pooling is a downsampling operation, which reduces the number of parameters from the volume. Here, max pooling operation is implemented with the help of 2x2 filters which decreases the volume size to 16x16x16. Similarly, the second convolutional layer is obtained by using the kernel of 3x3 with 12 activation maps and max-pooling with 2x2 filters. The resultant volume is reduced to 8x8x12. The max pooling operation with filter 2x2 is implemented to create an embedding layer of size 4x4x12 (192 values). Embedding layer holds the reduced and encoded form of the input data, which is then used to reconstruct the input through a group of convolutional and upsampling layers. In fig. 2, it can be seen that decoder is just the mirror view of the encoder architecture which differs only be upsampling operation instead of max pooling. The loss between the input and the reconstructed output is calculated with every iteration of the model run. The training is stopped after the loss is stabilized. Fig. 3 shows the input and reconstructed output of the image tiles.

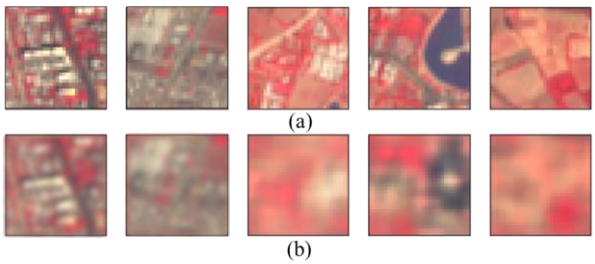

Figure 3: Input (a) and Reconstructed images (b) from the Autoencoder.

The CAE model is trained on a workstation with Multicore Xeon Processor with 32 Gigabytes of memory and 2 Gigabytes of Nvidia K2000 Graphics memory. The model is trained for two weeks until the loss subsided. The embedding values (192-D) generated by CAE are converted to 2D scatter plots with t-SNE for better visualization.

4. T-SNE CLUSTERING

t-distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear dimensionality reduction technique which preserves the relationship between data points while converting higher to the lower dimensions. t-SNE is used to find the relevant clusters by generating 2D representations of embeddings generated by CAE. The embedding values which are similar in structure form clusters. Embedding generated from the dataset lies in the data space R^D, where D = 192. The final representation of the data point after the implementation of the t-SNE algorithm can be given as R^2, where each data point is represented by a map point in the 2-D map space.

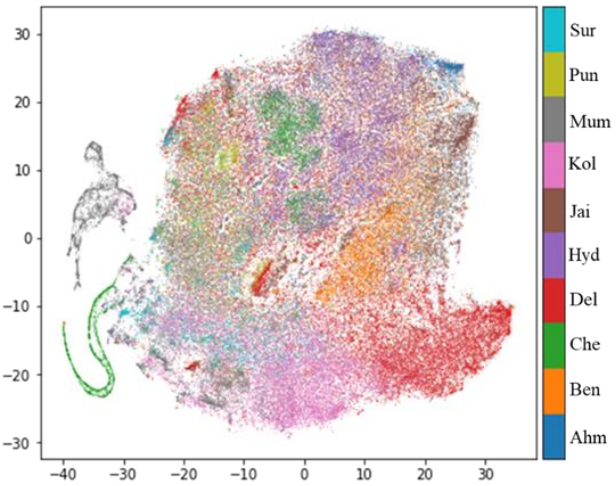

Figure 4: t-SNE representation of CAE embeddings.

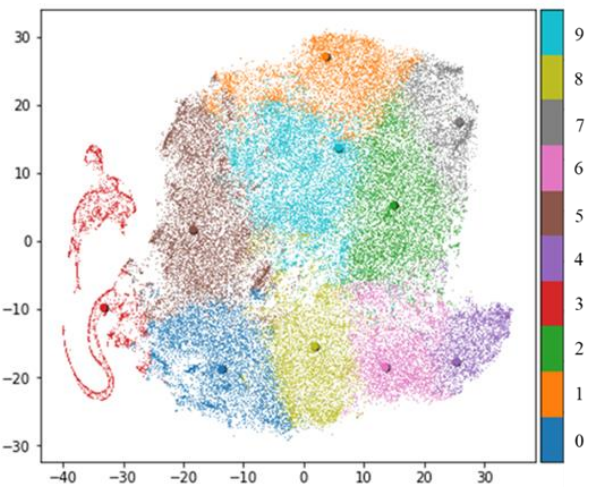

Figure 5: t-SNE representation with K-means clusters.

Fig. 4 shows the output of the t-SNE representation. t-SNE captures the local and global texture of the data space. The interesting patterns in the map representation can be delineated through different methods. (Tang et al., 2016) experimented with K-means algorithm on the High-dimensional dataset to delineate clusters in the dataset. This study utilizes the similar approach by implementing K-means clustering with k =10 to the t-SNE generated representation. The map points are clustered into 10 classes (Fig. 5). The clips corresponding to map points in each cluster are visually inspected to find general characteristics.

5. RESULTS

Table. 2 provides the general features of each of the created clusters which include various built forms, vegetation, water bodies, croplands, and forests. The distinction between the characteristics of each cluster is not perfect; however, by visual observation of the image tiles in each cluster, the general descriptions can be made.

Table 2: General characteristics of each Cluster.

Clusters	General description
0	Fallow lands, fewer croplands, less urban development, no visible road/railroads.
1	Open mid/high rise settlements, Moderate green cover, Open areas surrounded by buildings.
2	Fringe areas, scattered development, Visible road/railroads, Tree plantations, patches of crops
3	Sea, rivers, and lakes
4	Major croplands with minor patches of development especially roads.
5	A mix of compact settlements and Sparse built up, roads/railroads surrounded by urban settlements
6	Scattered low rise development like villages in urban boundaries, Tree plantations, croplands, roads/railroads surrounded by croplands.
7	Compact built form, less open spaces, fewer trees, uniform built typology.
8	Croplands, forests, water, almost no development, fewer visible road/railroads.
9	Compact, low rise development, fewer visible road/railroads, less vegetation, uniform urban fabric.

The clusters are studied in relation to the cities considered and the statistics regarding the presence of similar characteristics between different tiles is studied. Six among the ten clusters (1,2,5,6,7,9) show various urban built forms ranging from scattered built to densely compact and low-rise settlements. These clusters can be distinguished from the rest by presence of less vegetation and open areas among the built cover (Fig. 6). Cluster 1 includes the built-up areas with the presence of playgrounds and trees, while cluster 7 and 9 represents uniform urban fabric with less open and green cover. The city of Jaipur shows 42 percent of the area as low rise with dense built form, which might be due to the unique built typology prevalent in arid and semi-arid regions of the country. Cluster 0 includes the areas with open/exposed soil character. This feature can be attributed to the large playgrounds; area cleared of vegetation for development and naturally existing low vegetated area. The city of Pune and Ahmedabad shows the highest presence of cluster 0 at 9.8 and 7.2 percent respectively. The existence of rocky terrain and hills around the periphery of Pune city and the large parcels of land under development in Ahmedabad city can be one of the reasons.

Table 3: Percentage of image tiles of each city present in cluster 0-9.

Cities	0	1	2	3	4	5	6	7	8	9
Ahm	**7.2**	22.7	15.3	0.6	5.7	2.7	4.5	**32.0**	3.4	5.9
Ben	2.9	17.5	**22.1**	0.3	2.2	4.5	**10.5**	8.7	10.2	**21.1**
Che	0.6	9.6	15.6	**25.6**	4.0	10.7	5.5	15.0	2.4	11.0
Del	1.6	6.5	4.0	0.2	**16.6**	**15.5**	**12.0**	21.5	9.3	12.8
Hyd	0.2	**32.8**	**22.0**	2.0	0.3	5.0	0.3	13.2	0.8	**23.5**
Jai	2.1	17.9	**21.4**	0.1	1.7	1.8	1.7	**42.4**	2.6	8.2
Kol	5.1	10.0	4.8	3.5	**27.5**	11.2	3.8	9.0	**15.7**	9.5
Mum	4.6	5.1	2.8	**36.9**	2.1	5.1	2.1	11.3	8.9	**21.1**
Pun	**9.8**	19.9	5.7	1.0	1.1	**25.5**	1.2	21.3	2.9	11.7
Sur	2.4	**24.8**	10.7	5.7	10.4	5.1	4.4	16.3	**12.8**	7.2

The arrangement present in urban fringes includes scattered built form surrounded by croplands and transportation networks. Such characteristics are shown by cluster 2 and cluster 6, which further includes low rise development, exhibited by mainly villages in urban areas. The cities of Bengaluru, Hyderabad, and Jaipur show more than 20 percent of the patches in this category. Cluster 4 and 8 predominantly show croplands with varying levels of development. The boundaries considered in this study extend beyond the administrative limits which include the significant portion of city area as croplands. Due to this reason, the city of Delhi and Kolkata show large percentages of croplands (16 and 27 percent respectively). The water bodies are distinctly recognized compared to the rest of the clusters. The water bodies are present in the cities as lakes, rivers, and oceans. Mumbai and Chennai being coastal cities include large percentages (40 and 25 percent respectively) of the area as water Bodies.

6. LIMITATIONS AND CONCLUSIONS

This study presented a method to understand the inherent structure of ten Indian cities through unsupervised learning based on embeddings generated by CAE. This study experiments with the new image processing ideas and combine them with the task of classification of the urban landscapes. The study aimed to create a novel method to understand cities. However, there have been various shortcomings which can be solved in later studies based on this topic.

This study clips the patches of 32x32 pixels, which covers an area of approximately 10 Ha. The same methodology can be repeated with smaller clips for detailed urban studies. Tile size is a trade-off between the required details in the study and increasing computation costs. Further, the number of bands in the imagery can be increased from 4 to 13 to offer spectral variety to the CNN based model. The CNNs are sensitive to the tuning of hyperparameters, slight changes in these values may affect the quality of image reconstruction. The quality of current output may be significantly improved with minor tweaks in hyperparameter settings. The extent of the city considered in this study expands beyond the administrative boundary line, which provides a slightly amplified statistical figures in classification. Future studies may only consider the city area inside the municipal limits for better data comparison.

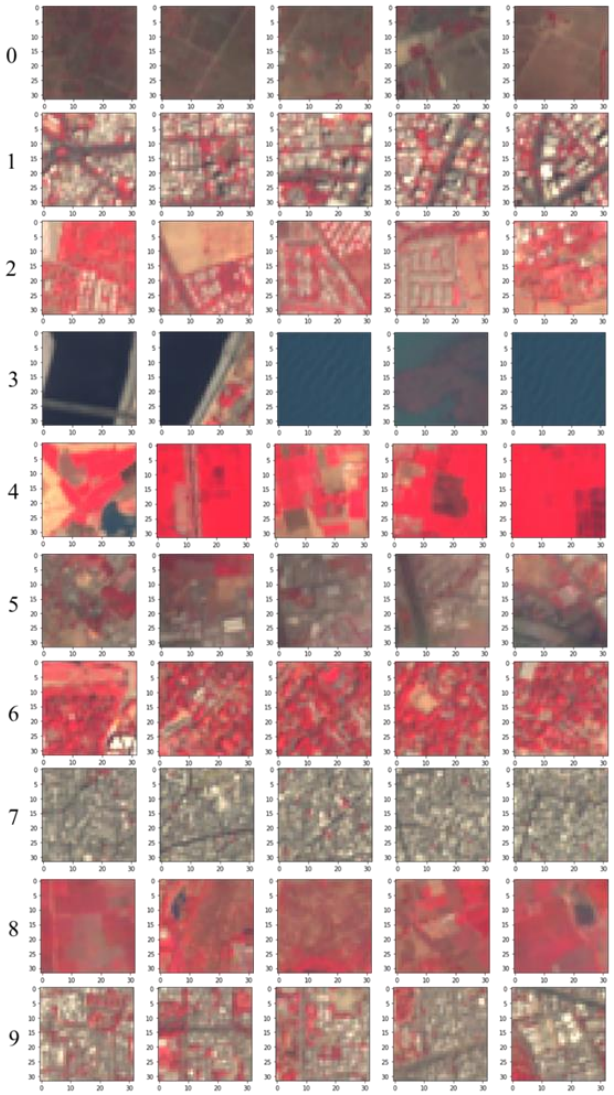

Figure 6: Example tiles present in each cluster.

Recently, studies have utilized Transfer learning approach which uses pre-trained Convolutional Neural Networks to produce embeddings, known as transfer values. Such pre-trained networks have been trained on multimillion images, which can find meaningful associations between extracted features in the images. Such methods can be applied to task demonstrated in this study, and the performance can be checked. Given the storage and large computational power, a large-scale study covering all the urban areas of the world can be combined and studied simultaneously with each other.

This study used the t-SNE algorithm extensively to plot the embeddings and to create clusters through nearest neighbors’ approach. However, the t-SNE algorithm does not preserve information regarding density and distances among the data points. The appropriateness of created clusters by K-means, which utilizes density and distance relation, is therefore questionable. The alternative options are the usage of self-organizing maps (SOM), which is an unsupervised learning method based on Artificial neural networks.

The study is part of an exploration of newer methods and the applicability in answering some of the questions in urban

mapping studies. Future research involving the latest algorithms and approaches would enhance the understanding of the topic.

ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Human Resource Development (MHRD), India and Industrial Research and Consultancy Centre (IRCC), IIT Bombay for funding this study under the grant titled Frontier Areas of Science and Technology (FAST), Centre of Excellence in Urban Science and Engineering (grant number 14MHRD005).

REFERENCES

Bechtel B, Alexander P, Böhner J, et al. (2015) Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. ISPRS International Journal of Geo-Information 4(1): 199–219. DOI: 10.3390/ijgi4010199.

de Moel H and Aerts JCJH (2011) Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates. Natural Hazards 58(1): 407–425. DOI: 10.1007/s11069-010-9675-6.

G. M, Ching J, See L, et al. (2015) An Introduction to the WUDAPT project. Proceedings of the ICUC9. Meteo France (February 2016): 6.

Khatami R and Mountrakis G (2012) Implications of classification of methodological decisions in flooding analysis from Hurricane Katrina. Remote Sensing 4(12): 3877–3891. DOI: 10.3390/rs4123877.

Kuffer M, Pfeffer K, Sliuzas R, et al. (2017) Capturing the Diversity of Deprived Areas with Image-Based Features: The Case of Mumbai. Remote Sensing 9(4): 384. DOI: 10.3390/rs9040384.

Mboga N, Persello C, Bergado JR, et al. (2017) Detection of informal settlements from VHR images using convolutional neural networks. Remote Sensing 9(11). DOI: 10.3390/rs9111106.

Noi PT and Kappas M (2018) Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. Sensors (Switzerland) 18(1). DOI: 10.3390/s18010018.

Shao Y and Lunetta RS (2012) Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS Journal of Photogrammetry and Remote Sensing 70. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS): 78–87. DOI: 10.1016/j.isprsjprs.2012.04.001.

Szostak M, Hawryło P and Piela D (2018) Using of Sentinel-2 images for automation of the forest succession detection. European Journal of Remote Sensing 51(1). Taylor & Francis: 142–149. DOI: 10.1080/22797254.2017.1412272.

Tang J, Liu J, Zhang M, et al. (2016) Visualizing Large-scale and High-dimensional Data. DOI: 10.1145/2872427.2883041.

Taubenböck H, Wegmann M, Roth A, et al. (2009) Urbanization in India – Spatiotemporal analysis using remote sensing data. Computers, Environment and Urban Systems 33(3): 179–188. DOI: 10.1016/j.compenvurbsys.2008.09.003.

Van Der Maaten LJP and Hinton GE (2008) Visualizing high-dimensional data using t-sne. Journal of Machine Learning Research 9: 2579–2605. DOI: 10.1007/s10479-011-0841-3.

Zheng B, Myint SW, Thenkabail PS, et al. (2015) A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. International Journal of Applied Earth Observation and Geoinformation 34(1). Elsevier B.V.: 103–112. DOI: 10.1016/j.jag.2014.07.002.