Spatial structure pattern prediction using deep learning technique for land cover modelling

DAMANDEEP SINGH
June, 2017

INDIAN INSTITUTE OF REMOTE SENSING
Indian Space Research Organisation
ISO 9001:2008

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION,
UNIVERSITY OF TWENTE,
ENSCHEDE, THE NETHERLANDS

Spatial structure pattern prediction using deep learning technique for land cover modelling

DAMANDEEP SINGH
Enschede, The Netherland, June, 2017

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geo-informatics

SUPERVISORS:
Prof.Dr.A.A.Voinov
Mr Ashutosh Kumar Jha

THESIS ASSESSMENT BOARD:
Prof.Dr.Ir.A.Stein
Prof.Rahul Dev Garg (IIT-R)

OBSERVERS:
Dr.S.Saran
Dr.V.A.Tolpekin

INDIAN INSTITUTE OF REMOTE SENSING
Indian Space Research Organisation
ISO 9001:2008

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION,
UNIVERSITY OF TWENTE,
ENSCHEDE, THE NETHERLANDS

DEDICATED TO MY PARENTS AND BROTHER

ABSTRACT

A land is defined as that portion of the earth which is not enveloped by water. The land cover takes into account the various objects present on the landscape such as agriculture, forest, barren land, built-up etc. Due to anthropogenic pressure, land cover is changing rapidly. If the land cover change is not monitored and controlled, soon the natural as well as the derived resources such as water, minerals, vegetation etc. will vanish. Therefore it is important to keep a check on the land cover change. It is not always possible to keep a track of every place for land cover change as some of the places are unexplored and inaccessible by humans. Remotely sensed images help in monitoring the land cover change by providing the repeatable observations of the same area and also of the areas that are inaccessible.

Modelling is given an aforethought for exploring the changes in land cover. Socio-economic and bio-physical factors affect the land cover change. Land cover change is modelled using the remotely sensed images and the factors that affect the changes such as population, temperature etc.

Machine learning models perform better than the other models such as cellular automata-markov (CA-Markov) in predicting the land cover changes. Machine learning algorithms, such as neural network has the ability to capture the non-linear relationship among the driving factors that causes the land cover change. Recent researches shows that deep learning neural network has the capability to model higher layers of abstraction in the data. There are several hidden layers in between the input and the output layer, which enables the deep neural network to extract the non-linear features from the data.

Since land cover change is a heterogeneous process that allows the interaction of the various factors such as bio-physical and socio-economic factors, deep learning helps in acknowledging the spatial pattern associated with the driving factors.

In this study, deep learning is used to model the land cover change and look for any spatial pattern associated with the drivers. The deep neural network is tested with the different cases of sensitivity analysis and also by varying the model parameters such as varying the number of iterations, the number of neurons in each hidden layer and it is found that it has the potential in land cover predictions.

Keywords: land cover change, modelling, deep learning, prediction, neural network

ACKNOWLEDGEMENT

The five months of intense research was an ambitious task. Finally after getting all the work done, thanking all those who were a part of this intense research.

I would first like to thank Guru ji, for his blessings are always with me. I am also thankful to my parents who stood beside me in every situation.

I am grateful to Dr. Senthil Kumar, Director, Indian Institute of Remote Sensing (IIRS) for awarding me this opportunity of pursuing IIRS-ITC joint education Masters of Science (M.Sc.) program 2015-2017

I am thankful to Dr.S.K.Srivastav of IGBP team for providing me the data for carrying out this research.

I am thankful to my ITC supervisor, Prof.Dr.A.A.Voinov for his expert guidance.

I am immensely thankful to my IIRS supervisor, Mr.Ashutosh Kumar Jha, Scientist/Engineer-SD, GID department for his constant support.

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Introduction

Agile and unplanned urban growth in large areas over the world constitutes of problems ranging from unsustainable development to environmental degradations (United Nations, 2014). With the change in the trend of the urban environment, land cover information also changes (Dewan & Yamaguchi, 2009). Problems caused by this change in the land cover, require decisions and responsive plans from various authoritative bodies such as government agencies, economy planners and policy makers to develop and implement relevant land policies. Timely updates and regular reports about the land cover are required to understand the changes in the classes that occurs over time.

Understanding the spatiotemporal behaviour of land cover is important for analysing the trends in the spatial structure of the classes (X. Zhang, Kang, Wang, & Sun, 2010). Remotely sensed satellite images provide repeatable observation about the land cover; thus becoming an important mean to acquire land change information. After the acquisition and classification of the images, these images can be used to map the changes such as conversion of agricultural area into urban. For modelling the spatiotemporal characteristic of land cover, it is important to consider the revisit time of the satellite acquiring the image.

To date, several methods have been tried out to model the land change. The choice of a method primarily depends on the user's interest and the availability of datasets. Various modelling methods have been used to model land cover change, such as modelling spatiotemporal model of land use change based on ant colony optimisation (Yang, Zheng, & Lv, 2012), simulating urban expansion by integrating Markov chain, cellular automata and logistic regression (Arsanjani, Helbich, Kainz, & Boloorani, 2012), agent-based modelling (Vermeiren, Vanmaercke, & Van Rompaey, 2016).

Several modelling techniques in remote sensing, already exist such as the well-known cellular automata (CA), Markov chain, integration of the CA-Markov and other approaches. The problem with the CA is that it gives too much emphasis on the local interaction of pixels rather than on interpreting the spatiotemporal patterns. Markov chain is not useful to estimate the position of change, it controls temporal dynamics among the land cover types through the use of transition probabilities i.e. it gives the probabilities of a particular class to be converted into other class, and therefore, it needs to be combined with auxiliary spatial models. Modelling by using CA-Markov is also not preferred because the simulated map contains the salt and pepper noise i.e. white and black dots on the image, this is because it relies on the conditional probability images produced by Markov to inhibit the spatial distribution (Yang et al.,

2012). Moreover, it does not take into account the socio-economic factors such as owner's willingness to convert a land.

Land cover can also be modelled using UrbanSim. Although it also has several limitations such as economic interactions are not taken into consideration. Moreover, UrbanSim requires a plentiful number of packages such as Numpy, Scipy, Enthought python (Kakaraparthi & Kockelman, 2011).

Spatial dynamics of land change can be modelled by using CLUE model. The limitation of using this model is its inefficacy to handle mixed pixel such as pixel that is partially urban and partially agriculture, as opposed to categorical data for which it performs well (Veldkamp & Fresco, 1996). All the models mentioned above are based on idiosyncratic defined rules. These rules are based on the individual's experience of the study area and the subject and therefore it incorporates some bias towards the assumptions made (Maithani, 2009).

There is need of effective modelling technique that is able to simulate the changes without incorporating the effects of noise, dependency on vague assumptions, and getting ambiguous results. It has been observed that neural networks perform well when tackling with poor data and catching non-linear complicated features in modelling process. Neural networks appear to be most reliable for simulating complicated relationships (Li & Yeh, 2002).

The deep neural network is differentiated from the general single-layer hidden network with respect to their depth i.e. the number of multiple hidden layers through which the data passes. Deep neural network is capable of recognising the nonlinear complicated features in the dataset. In this research work, the deep neural network is explored for predicting the spatial structure of the various classes for better understanding of the land change environment based on land change drivers. This will assist the researchers and government agencies in taking an important decision about land change to formulate land policies.

1.2. Research Identification:

A lot of modelling techniques have been developed, to understand the trends or patterns in the land cover change incorporating different techniques such as regression techniques, CA-Markov and machine learning algorithms. This research coils around the development, analysis and assessment of the deep learning technique for the prediction of the spatial structure i.e. grouping or association of the different classes with the other classes.

1.3. Research Objectives:

The main objective of this research work is to predict the spatial structure to explain the changes in the land cover using deep learning algorithms, for improving the prediction results. The main objective can be reached by defining the following sub-objectives:

1. Study the land change patterns by analysing the spatial structure.

2. Evaluate the capabilities of deep learning algorithms for predicting the spatial structure from satellite images.

3. Identify the drivers responsible for the land cover change.

1.4. Research Questions:

The following are the research questions identified from the research objectives:

1. What are the drivers or the factors that affect the land change over time?

2. How will the spatial pattern such as areas that are adjacent or nearby to one another, lead to the prediction of spatial structure?

3. What is the relation between spatial structure and land change drivers?

1.5. Innovations Aimed At:

The innovation aimed by the research work is:
Only limited research has been carried out with deep learning algorithms with respect to remote sensing images as most of the practical applications are found in a more general context such as in computer vision and language domains. Therefore this research explores the capabilities of deep learning algorithms for predicting the spatial structure to explain the trends or pattern for the land cover change. In this research, efforts are made to recognise the spatial pattern in the images. On the basis of the recognised pattern and statistical measures, prediction can be made.

1.6. Thesis Structure:

The research is organised as follows:

Chapter1: Introduction, research objectives and research questions considered in this work.

Chapter2: Literature review on land change modelling techniques and deep learning

Chapter3: Study area and the data set used in this research work

Chapter4: Methodology adopted for carrying out the research.

Chapter5: Results and discussion about the results obtained from the model.

Chapter6: Conclusion and recommendations of the research work.

2. LITERATURE REVIEW

2.1. Land cover change and its impact

The work by Newman, (2006) shows that as the city expands, the cost of infrastructure and housing grows; since the availability of land, water, and other resources are limited, the problem of congestion occurs. Therefore, it has become a matter of concern to study the process of land cover change.

Land cover is considered a crucial aspect in geographical as well as in geological analysis. It represents the interactions between bio-physical and socio-economic factors or drivers. Land cover features are complicated as an outcome of the increasing population pressure and human activities (Rujoiu-Mare & Mihai, 2016). Trend and pattern scheme of the land cover change is considered crucial for managing natural resources. Modelling is given aforethought to explore the changes in the land cover. Modelling land change is a heterogeneous process that involves a number of factors responsible for the change such as the socio-economic and the biophysical factors. These factors help in acknowledging the spatial patterns and evaluating the land cover change (Alvarez Martinez, Suarez-Seoane, & De Luis Calabuig, 2011)

Analysing the spatial patterns is considered mandatory for a perspective view of the complex spatial mechanism which is responsible for the occurrence of phenomena. Modelling the spatial distribution requires the essential morphology of the spatial pattern (Chou, 1995). The spatial structure of the patterns is described as the grouping of the singular entity distributed over an area and the topographic relation between them.

Monitoring and analysing the changes in the spatial structure of the classes that constitute the land cover is essential for growth, development, and validation of planning strategies. But monitoring these changes at the national level is not an easy task. This is why remote sensing is considered as a powerful tool to monitor changes in the spatial structure and to look for the changes in the land cover (Schneider, 2012). With the arrival of LISS and LANDSAT remote-sensing satellites, mapping land cover changes is possible at variable scales. Mapping these changes at multiple temporal and spatial resolutions helps in land cover prediction modelling. The remote sensing satellite provides the historical and the present information that is essential for observing and analysing the changes in landscape drivers or factors. (Sohl & Sleeter, 2012)

As the effects of the land cover change on bio-diversity and climate have been observed, modelling these changes is necessary and meaningful (Zessner et al., 2016). Various researches have been conducted in the past that shows how important it is to model the land cover changes. For example, Brink et al., (2014) have discussed, how the interaction of the anthropogenic activities with the environment effects the landscape and the biodiversity at spatial and the temporal scales. Therefore monitoring and analysing the changes due to human activity on the land cover, is important for managing natural resources.

2.2. Different types of model for land cover change

There are several approaches to model the land cover change such as:

- **Equation-based models**: The models are based on some mathematical equations. For example, an equation linking the population with the area change over time.
- **System Models**: The model is based on stock and flow approach. This type of model incorporates the interaction between the humans and environment.
- **Statistical Models**: It is one of the most used method for modelling the land cover change. It includes the different regression methods. Statistical techniques are proven effective unless they are used with theoretical models
- **Expert Models**: This model incorporates the combined effect of expert knowledge and the probability functions such as bayesian probability.
- **Cellular models**: It includes the Markov, Cellular Models (CA) or a combination of the two.
- **Agent-Based Models**: It incorporates the human decision making along with their interaction with the environment (Parker, Manson, Janssen, Hoffmann, & Deadman, 2002).

As mentioned above, many techniques have already been developed for modelling the land cover change. However, each of the above technique used has one or the other limitation. For example, equation based models are difficult to simplify in order to reach to an analytical conclusion. The Kernel-based approach has been used as a learning technique to accommodate the non-linear transition rules (CA) for simulating land cover change. But it also has the limitation of not taking into account the morphology of spatial structures, such as connectivity (Liu et al., 2008).

2.3. Machine learning models

When modelling the nonlinear complex features of land cover change, it has been observed that machine learning algorithms give better results than the other mentioned models (LeCun, Bengio, & Hinton, 2015). Machine learning models equip the machine or computer with the skills to be trained without specifically writing hundreds of lines of programming code. These models provide a kind of artificial intelligence to the machine or computers to look for patterns in the data and alter its decision

accordingly. Like classification algorithms, machine learning is also categorised as supervised and unsupervised learning. In supervised machine learning algorithms, target data is used for training the machine and get the required output while unsupervised machine learning draws conclusion from the input datasets (TechTarget, 2016). In a more general way, machine learning is another approach of analysing the data that automates the process of modelling. Machine learning is a modern approach to solve the complicated mathematical computations of big data, repeatedly and faster. Machine learning models such as artificial neural network, Random forest classifier, support vector machines and much more are already being used for land cover mapping. Gounaridis et al., (2016) research shows that random forest performs well over a wide area with a large number of heterogeneous classes.

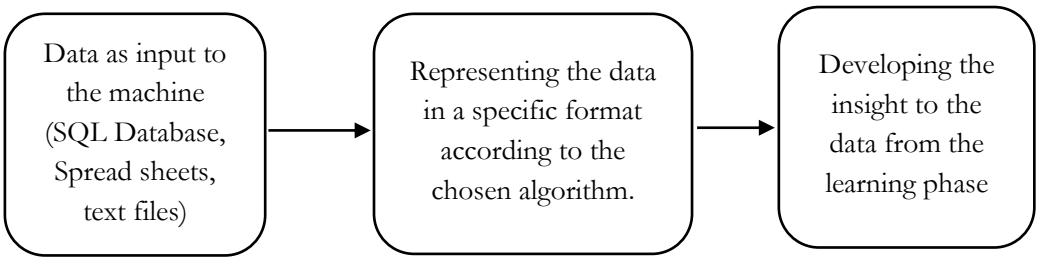


Figure 2.1: Workflow mechanism of machine learning models

2.4. Neural Networks

The neural network is a subset of machine learning and is inspired by the perception of the functioning of the human brain. In a more general way, it is defined as the system that models the procedure in which biological neurons execute a peculiar task. The neurons are the fundamental building blocks of the neural network and it acts as the data processing unit that receives input from the other interconnected neurons. The block diagram shown in figure 2.2 is that of a model of a neuron. Three basic elements are identified from the block diagram that shows the architecture of the neuron (Haykin, 1999)

- **Synapses**: It is the connecting link between two nodes and is characterised by its own strength or weight. These interconnected neurons drive the elemental process of how the network learn (Du & Swamy, 2014). Learning happens when the network adjusts the weight of the neurons according to the desired output.
- **Summation or adder (∑)**: It adds the product of the input units with their respected weights. As shown in the figure2.2 input unit X1 is multiplied with the weight W1. Similarly, all the input unit are multiplied with their weights. Summation is used for adding these products.

- **Activation function**: It is used for limiting the output of the neuron i.e. it squashes the output within a particular range. For example: range of tanh activation function is [-1, 1].

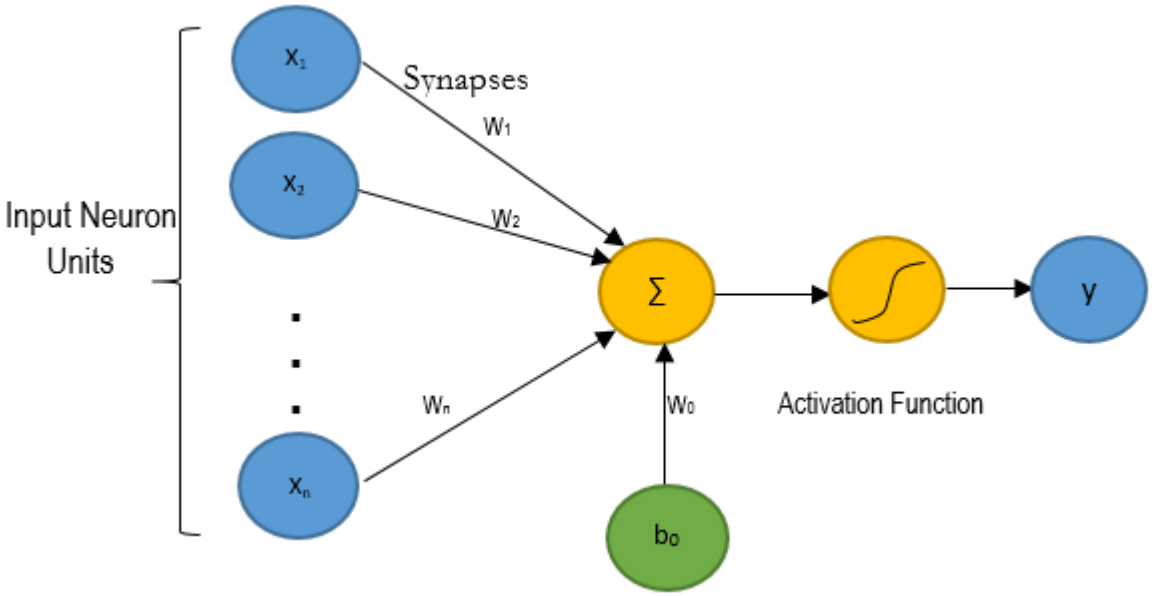


Figure 2.2: Architecture of a neuron

The output y is computed with the help of a mathematical equation that is formulated as below:

$$y = b0 + \sum_{n=1}^{n} Xn * Wn$$

2.5. Deep Learning

Deep learning is a subset of machine learning algorithms that allude to multi-layer interconnected neural network and has the capability to extract the spatial and the spectral information from the data (Nogueira, Penatti, & Jefersson, 2016). These multiple layers form a hierarchical structure of learning the high-level feature from the lower level.

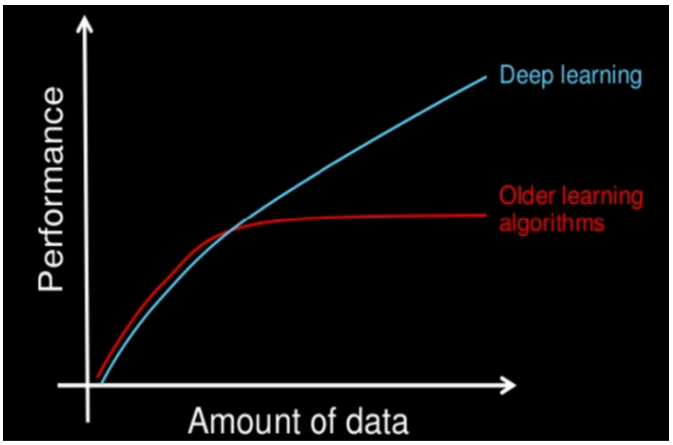


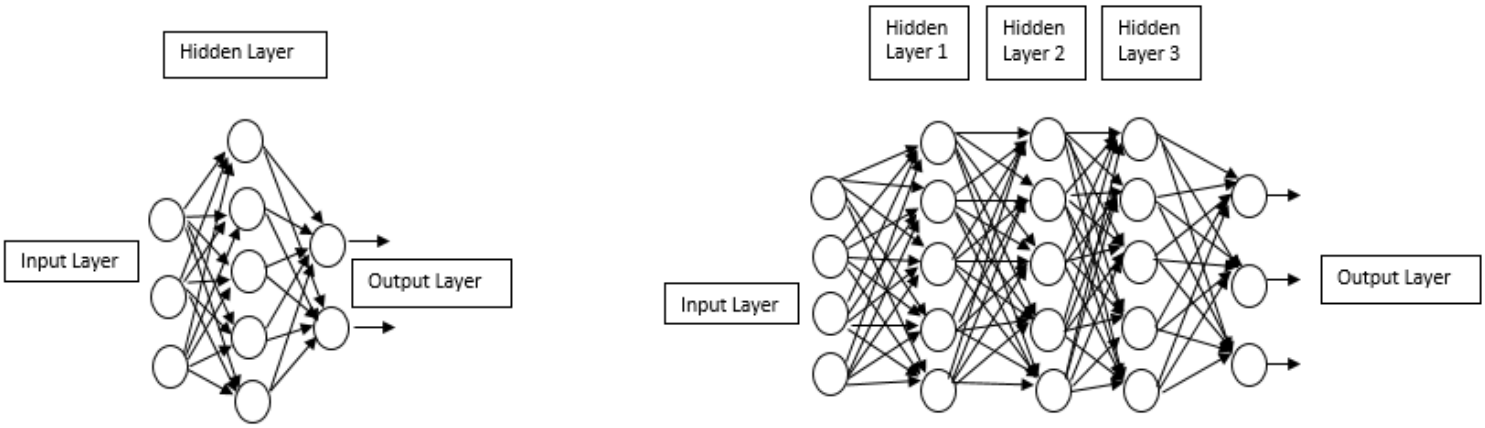
Figure 2.3: Performance graph of deep learning

Source: (Andrew Ng, 2015)

It is widely used because of its ability to retrieve different levels of abstraction ranging from detecting edges and corners to more semantic information in the initial layers, recognising shapes in the intermediate layers and high level information i.e. whole object in the final layer (L. Zhang, Xia, Wu, Lin, & Tai, 2015). It has been found that more precise prediction can be modelled when encoding the spatial associations between joint locations (Tompson, Jain, LeCun, & Bregler, 2014).

Based on the performance of the deep learning in machine learning domain, deep learning is also practised with remote sensing images such as the extraction of spectral and spatial information for classifying the images(Zhao & Du, 2016) , identifying cars from large point cloud data using deep learning (Yu, Li, Guan, & Wang, 2016).

2.5.1. Architecture of Deep Neural Network



'Non-deep' feed forward Neural Network

Deep feedforward Neural Network

The word architecture points out the overall structure of the network i.e. what type of activation function is used, how many neurons are used per layer, what is the width of the network.

The first question that arises, when dealt with "Deep Learning" is, why is deep learning called "deep" and how does it differ from the artificial neural network.

The "deep" in the deep learning refers to the depth or the number of hidden layers used for creating the neural network (Raut, 2017). For example, according to the research conducted by Goodfellow, Bengio, & Courville (2016) as shown in figure 2.3 experimental results have shown that deep networks generalize better when they are used for interpreting the digits from the photographs.

Before deep learning came into picture, artificial neural network was used. In general, artificial neural networks were created using only one hidden layer because of the limitation of the efficient hardware and software. The main issue with the artificial neural network was the problem of "vanishing

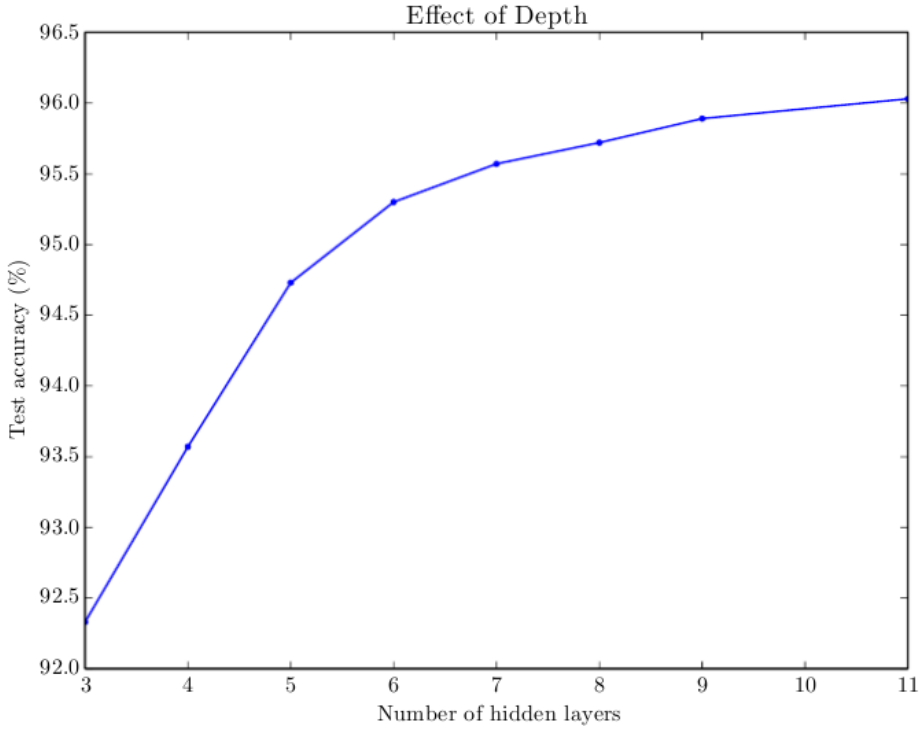


Figure 2.4: Effect of the number of hidden layers on accuracy of the experiment used for interpreting the digits from the photographs

gradient". According to Goodfellow et al., (2016), the problem of vanishing gradient means that it is difficult to adjust the error that occurs due to the difference between the target and the observed value, to layers that are closer to input units.

Merely increasing the number of hidden does not guarantee, better results. Empirical results have shown that deep neural networks perform well with RELU (Rectified linear unit) activation function because it is easy to optimise (Goodfellow et al., 2016).

Moreover, to increase the performance of the neural network, optimisation of the network is required. Optimisation techniques such as back propagation algorithm and gradient descent algorithm are used. Back propagation is a simple method of training the neural network and is used in combination with gradient descent algorithm. The output of the neural network is compared with the target and the difference between the two is calculated using the loss function. This difference is propagated backward in the network from the output layer to the input layer. Back propagation uses this difference to update the weight by determining the gradient of the loss function.

Deep learning algorithms rely on the machine learning library for all of its numerical computations and handling big data. Many libraries have already been used for creating the deep neural network.

This research uses a machine learning library named "Tensorflow". A brief introduction and advantages of tensorflow is described below.

Tensorflow: An open source machine learning library developed by Google for performing complex arithmetic computations. It uses tensor (multi-dimensional array) to communicate between nodes. It has a user-friendly interface for python. The library is still in its development stage and the researchers at Google are continuously updating the library and improving its performance.

This research focuses on the prediction of land cover change map involving deep learning using Tensorflow. Tensorflow supports different platforms such as tablets, mobile instruments, CPU, and GPU. It is more flexible in terms of programming environment than DistBelief (machine learning library developed by Google). Tensorflow supports several operations that are bind into its libraries such as MatMul (matrix multiplication), ReLU (rectified linear unit) and many more. It is tested for local (single machine) as well as for distributed systems (GoogleResearch, 2015).

A deep neural network using tensor flow was used for classifying the handwritten digits achieving an accuracy of 99.2%. The model consisted of two convolutional layers and uses dropout method (a technique for reducing the over-fitting of data) (GoogleResearch, 2015).

InceptionV3 a model developed using tensor flow for ImageNet (an image database consisting of millions of images) that learns to classify the images into 1000 category like "leopard", "container ship" has an error rate of 3.46% whereas AlexNet achieved an error rate of 15.3% and BN-Inception –v2 achieved error rate of 4.9% (GoogleResearch, 2015).

2.6. Type of Tensorflow Models

Tensor flow is used for training on large datasets. It has the ability to simultaneously train the deep feed forward neural network and wide linear networks

- **Wide linear network**: It offers the benefits of Memorisation. Memorisation can be defined as the periodic appearance of features and correlate it with the past data.

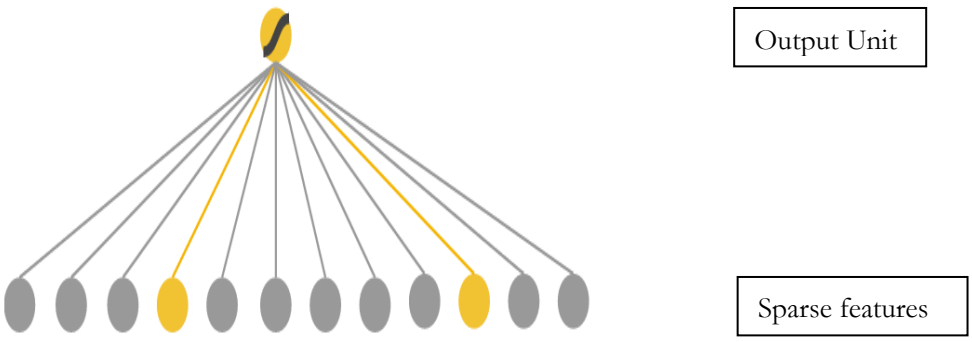


Figure 2.5: Tensorflow wide model

- **Deep Models**: It offers the benefits of generalisation. Generalisation refers to the finding of new feature combination that has occasionally appeared in the past.

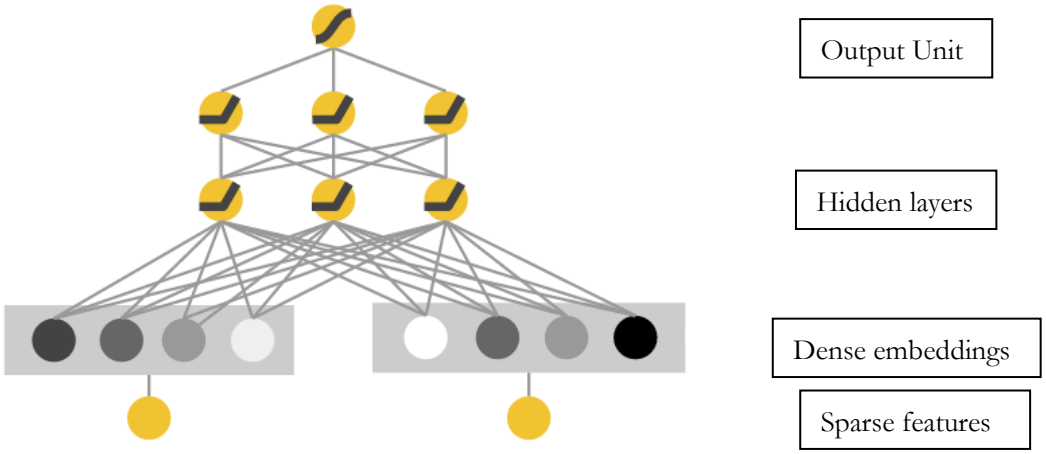


Figure 2.6: Tensorflow deep model

- **Wide and deep models:** It combines the capability of both the wide and deep models i.e. memorisation and generalisation. (GoogleResearch, 2015)

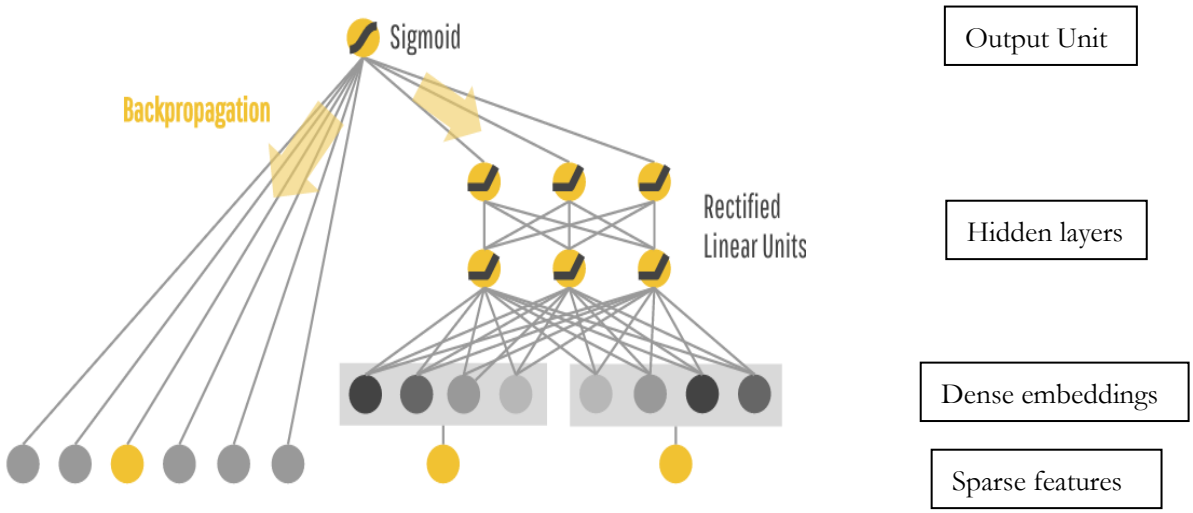


Figure 2.7: Tensorflow wide and deep models

3. STUDY AREA AND DATASET USED:

3.1. Study area

India is a part of South Asia and is known for its diversity in physical features, flora and fauna, river systems such as mountains, tropical forest, rivers etc. Climate, vegetation, population, natural resources vary from place to place. The river system in India is classified into 1) Himalayan rivers 2) Peninsular rivers (Prafull, 2015). Ganga is one such river that comes under Himalayan river belt. The rivers play a vital role for the people of India as it satisfies the basic need of water for various purposes such as water used for drinking purposes. Moreover, some of the rivers in India are considered sacred. Ganga basin is considered sacred and is situated in northern India that covers 981,371 square kilometers. At Haridwar, a place at the base of Himalayas, Ganga is divided into Upper Ganga Canal. This research investigates the Upper Ganga Basin for land cover change.

The river Ganga emerges from Gangotri glacier as Bhagirathi and is accompanied at Devprayag by Alaknanda. The river Ganga flows through the region of Bihar, Uttar Pradesh, West Bengal and parts of Uttarakhand. The study area of the research is Upper Ganga Basin (UGB) that has a stretch of 87,787 km2. The UGB measures moderate rainfall in the range of 550-2500mm (Bharati, L. and Jayakody, 2010). The altitude in the UGB varies in the range of 7500m-100m. The river Ganga is currently suffering from high pollution levels and is also affecting the life of the communities living nearby. Waste from the industries that are set up along the river pollutes it. Human activities such as religious offering in the form non-biodegradable polymer add larger volumes of pollutants to the river. As the population of India continues to grow, the number of peoples visiting the temples and shrines of the Ganga basin has increased. Tourist act as a source of pollutants that are being added to the river. This in return puts a pressure on the government body to maintain the sanctity of the river. Moreover, the population of the Ganga basin has also increased in the past. Therefore it has become a matter of concern to study the patterns and trends of the land cover change in UGB.

According to the technical report of Indian Space Research Organisation (ISRO) Geosphere Biosphere Project (IGBP), any change or transition in the river basin is a result of the land cover change. Due to this change, various processes such as bio-diversity, soil erosion is affected and this in return affects the human habitability. Therefore estimating the impacts of the land cover change is considered crucial by modelling the relationship between the driving factor and the land cover change classes such as built-up, crop land etc. (Use, Cover, & Change, 2015)

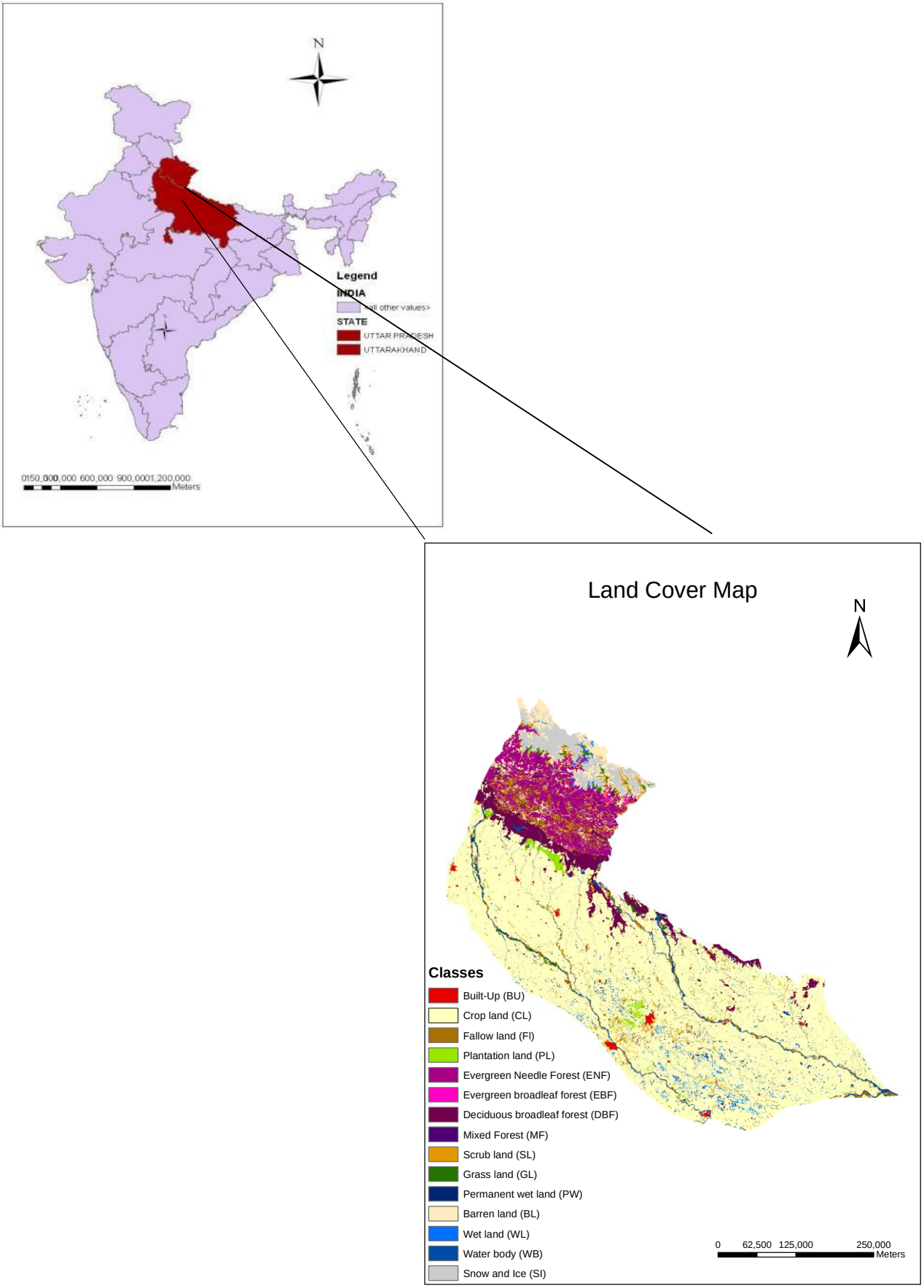


Figure 3.1: Upper Ganga Basin

3.2. Land cover change and driver datasets:

3.2.1. Land Change drivers or factors:

Changes in the land cover is a result of multiple factors occurring over the concerned study area. These driving factors can be categorised into two categories:

- **Driving factors that arises due to human activity**: Among the anthropogenic activities, the increase in population plays a major role, but it is not the only factor that causes the land cover change. Construction of the new roads leads to settlements nearby. It is observed that wherever the road network density is high, built-up areas are observed. The driving factors due to anthropogenic activity can be considered under socio-economic factors. Following socio-economic factors are considered:

1. **Population density**: The demographic data is obtained from the Office of the Registrar General and Census Commissioner, which comes under Ministry of Home Affairs, India. Using the census data, the population density map of three years 1985, 1995, 2005 is prepared at a scale of 1:250000.

2. **Road Network**: Roads are considered crucial for the overall development of any area. Settlements or built-up grows along the road and its periphery. The road network map is prepared for the study area using the toposheet of Survey of India.

- **Driving factors originating due to natural phenomena**: Natural factors also accounts for the change in land cover. For example climate is the utmost natural factor that effects the land cover from yearly to decadal time ratio. The changes in the climate effects the crop field which in turn affects the soil. Rain also effects the soil, as in cases of heavy rain the top fertile layer of the soil is washed away, leaving the soil unfertile. The driving factor due to natural forces can be considered under the bio-physical factors. Following bio-physical factors are considered:

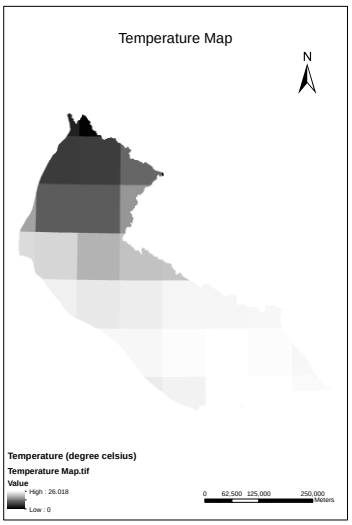
1. **Soil depth**: Soil map is prepared by the Soil and Land Use Survey of India. The soil map is prepared for the upper ganga basin that represents the depth of soil in various regions of the study area. The map is prepared at a scale of 1:1million scale.

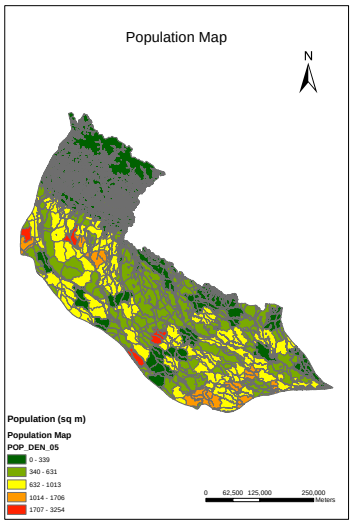
2. **Temperature**: Indian Meteorological Department is the governing body that maintains the temperature data. The temperature map is prepared at 0.5 degree resolution using shepard interpolation technique (A. Srivastava, Rajeevan, & Kshirsagar, 2008). The yearly mean temperature is calculated for 2005, 1995 and 1985.

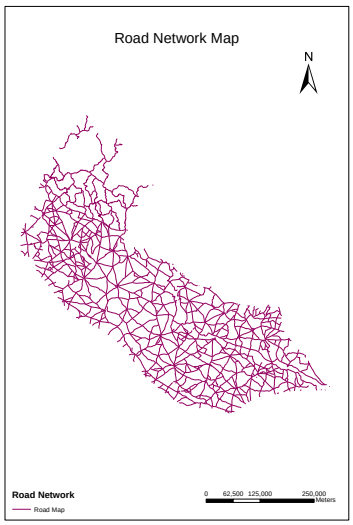
3. **Rainfall**: Indian Meteorological Department prepares the rainfall statistics. The rainfall map is prepared at .25 degree resolution for the three time series map i.e. 2005, 1995 and 1985. The total rainfall for the year is calculated by adding the rain of each individual day for the particular year (Pai et al., 2014)

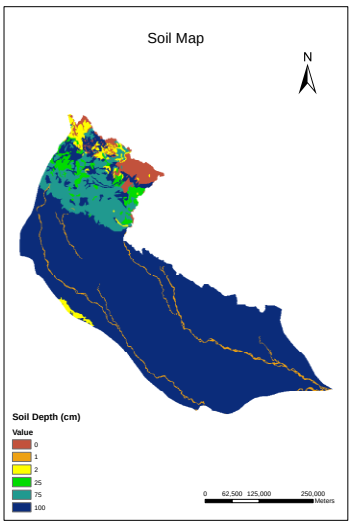
4. **SRTM**: The SRTM map contains the elevation data. The map is prepared at a resolution of 90m. For effective modelling of the land cover change, SRTM data plays a vital role. SRTM is same for all the three years because it remained constant over the time period.

3.3. Driver Datasets:









The land cover map for the Upper Ganga Basin is obtained from the Indian Space Research Organisation (ISRO) Geosphere Biosphere project (IGBP). The map is classified in conformity with the classification scheme of IGBP project. Land cover maps of years 1985, 1995, 2005 are generated at a scale of 1:250000 using satellite data such as:

- LANDSAT MSS(1985)
- LISS I (1995)
- LISS III (2005)

The map shown below is the classified land cover map of the study area. The map is classified into fifteen different classes according to the classification scheme of IGBP. The map is classified into:

1. Built-up (BU)
2. Crop land (CL)
3. Fallow Land (FL)
4. Plantation land (PL)
5. Evergreen Needle Forest (ENF)
6. Evergreen broad leaf Forest (EBF)
7. Deciduous Broad Leaf Forest (DBF)
8. Mixed Forest (MF)
9. Scrub Land (SL)
10. Grass Land (GL)
11. Permanent Wet Land (PW)
12. Barren Land (BL)
13. Wet Land (WL)
14. Water Body (WB)
15. Snow and Ice (SI)

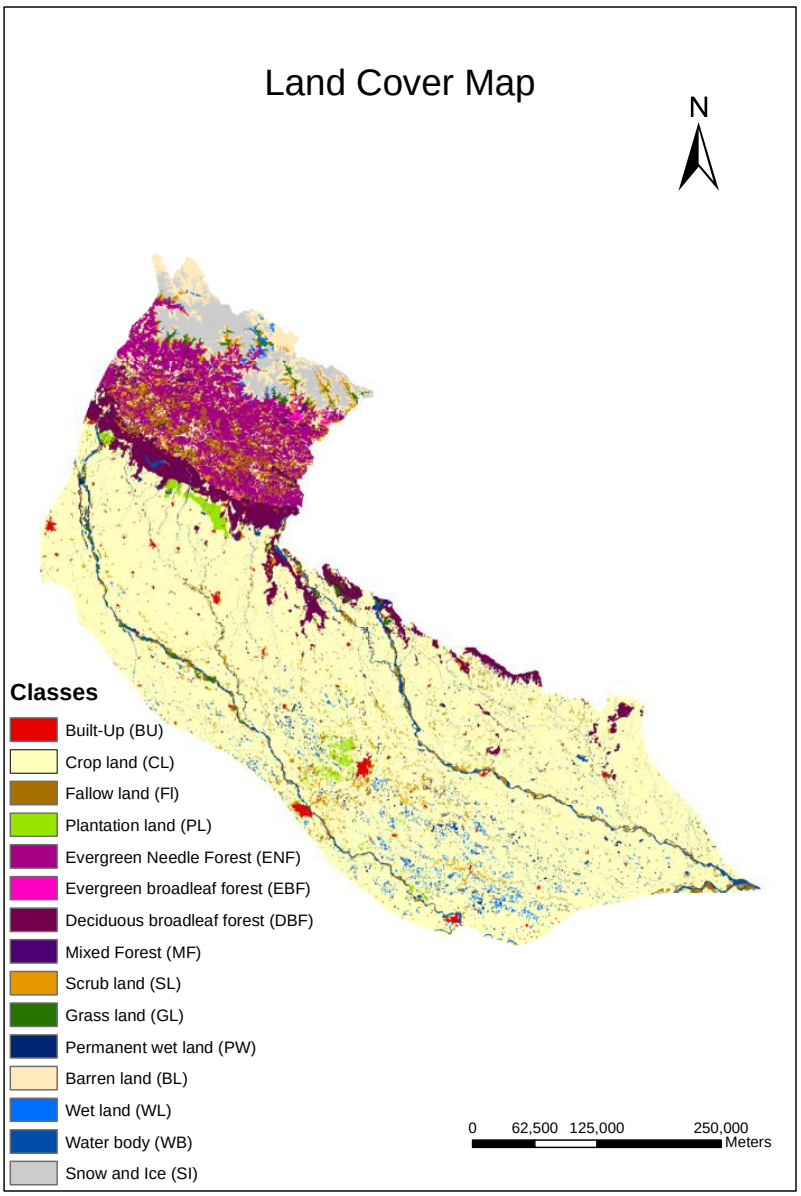


Figure 3.2: Land Cover Map

4. METHODOLOGY

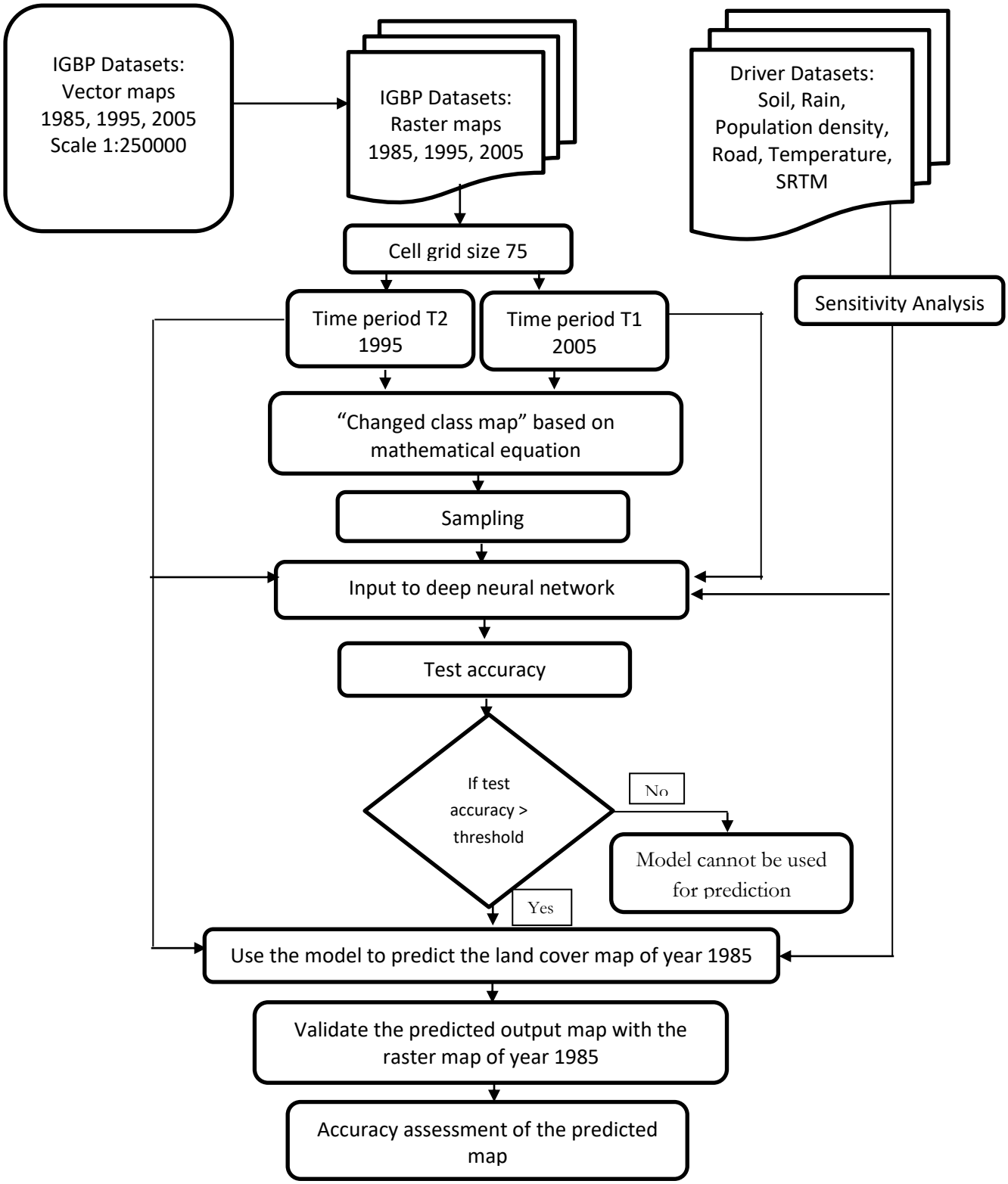


Figure 4.1: Methodology

4.1. Description of methodology:

1. The vector maps prepared by the Indian Space Research Organisation (ISRO) for its GBP project for the years 1985, 1995 and 2005 are prepared at a scale of 1:250000. The vector maps consist of various classes such as built-up, crop land and other classes as mentioned in chapter three, in the form of polygons. The raster maps i.e. the maps in the pixelated form are prepared from the vector maps at a cell grid size of 75m.

2. In this research, back casting or backward prediction is performed. It is assumed that the drivers of the year 2005 are responsible for the changes in the land cover map of the year 1995. The road map, which is one of the drivers for predicting land cover change is available for the year 2005 only. Since road map is available for the year 2005 only, using this road map, the backward prediction or back casting is done for the year 1995 for performing the accuracy assessment of model. This allows the neural network to establish the relationship between the land cover maps of the year 2005 and 1995 and its corresponding drivers. This model is then further used for predicting the land cover map of the year 1985.

3. For predicting the land cover map of the year 1985, a deep neural network model is designed by using the land cover map of the year 2005 and 1995 and driver dataset. In this research, two approaches have been used to train the neural network. In the first approach, an equal number samples are taken from each of the classes by creating a map called "changed class map" from the land cover maps of the year 1995 and 2005 and is given as input to the deep neural network. In the second approach, the whole land cover maps of the year 1995 and 2005 are directly given as input to the deep neural network along with driver datasets. Both the approaches are explained below:

   First approach (Sampling performed)

   Consider the year 2005 as time period T1 and year 1995 as time period T2. A mathematical relationship is developed between the two land cover classes in order to make the deep neural network learn the changes for the two time periods. The neural network is designed in such a way that it learns the features from both the time period and takes the effective decision to recognise the classes which are converted from one class to another and also learn the classes which remained same between the two time periods.

   The mathematical equation relating the different classes of the two time periods i.e. T1 and T2 is formed as:

   Changed class map = (class in T1*100) + class in T2

For example a pixel is labelled as class 10 in time period T1 at some location (x, y). The same pixel is labelled as class 11 in time period T2 at the same location as that of time period T1. According to the mathematical equation stated above:

Changed class map = (10*100) + 11
Changed class map = (1000) + 11
Changed class map = 1011
The first two digits of the changed class map represent the class 10 i.e. in time period T1 the pixel belongs to class 10 and the last two digits state that the class 10 gets converted to class 11 in time period T2.
If in the above example, class 10 in time period T1 at some location (x, y) remains in class 10 in time period T2 at the same location as that of T1. Then according to the above mathematical equation:
Changed class map =1010

This changed class map represents the number of pixels of the classes where the changes have taken place and the places where the classes remained same in time period T1 and T2.

Random sampling is performed on the changed class map in order to take an equal number of samples from both the maps where the changes have taken place and the places where the classes remained the same. Hence equal conversion probability of the all the classes are considered. This results in unbiased sampling.

Second Approach (complete map are given as input to deep neural network)
In this approach, the complete land cover maps of both the time periods i.e. 2005 and 1995 are given as input to the deep neural network along with the driver datasets. This approach is used considering the fact that the distribution of the various classes in the land cover maps is not uniform. For example: in all the land cover maps of years 1985, 1995 and 2005, the crop land is dominant over all the classes. This allows the neural network to adjust weights according to the actual distribution of the various classes.

4. In both the approaches explained above, the land cover maps of the two time periods i.e. 2005 and 1995 along with the driver datasets of the year 2005 are given as an input to the deep neural network for the network to learn from the labelled datasets. It is assumed that the drivers of the year 2005 are responsible for the changes in the year 1995.

5. Apart from the maps that are given as input to the deep neural network, the tensorflow library allows the user to define the parameters of the deep neural network. The parameters that user can define are:

    - Hidden units: It defines the number of neurons per hidden layer
    - N_classes: It defines the number of the classes to be predicted
    - Optimiser: The model uses as Adagrad optimiser. According to Ruder (2016), it is observed that adagrad enhances the performance of the stochastic gradient descent and is used for training large neural networks.
    - Activation function: Rectified linear unit is used as an activation function.
    - Drop out: drop out is used as a regularisation technique. Dropout is a method in which some of the neurons are dropped out from the network i.e. temporary discarding some neurons and all of its connections with the other neurons. By using the drop out method, it is observed that the error reduces significantly in comparison to the L1 and L2 regularisation methods (N. Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014)

6. The factors or the drivers responsible for the land cover changes are prepared at different resolutions. For examples the temperature map is prepared at 0.5 degree resolution, rainfall map is prepared at 0.25 degree resolution.

7. For modelling the land cover changes, all the driver dataset maps and the land cover maps should be at the same cell grid size. Raster maps are prepared from the vector maps. Information present in the remote sensing images is always scale dependent. Therefore scale and cell grid size of the map plays a vital role in displaying the information of the classes such as the area covered by the different classes, the image is classified into and also for visualising the images. The cell grid size of the raster maps is calculated using the scale of the vector maps. Cell grid size comes out to be 75m. For effective modelling, all the raster maps of driver dataset used are of grid size 75m. Bringing all the datasets to same grid size is not sufficient, extent i.e. the number of rows and column, no data value, and other parameters should also match of all the datasets.

8. Deciding the drivers responsible for the land cover change is considered vital. As it is assumed that the changes in the land cover that occur are due to the drivers used in the modelling process. Various approaches are used for selecting the appropriate drivers such as running a logistic regression, overlaying one driver with the other and finding the relations between them or performing a sensitivity analysis.

9. Sensitivity Analysis:

Sensitivity analysis is an indicator that estimates the variation in the model output with respect to change in the model input parameters. It is also used for estimating the robustness of the model. According to Burg (2016), feeding a large amount of data into the model does not always result in better outputs. In fact, the more the data fed into the model, the higher the chances of introducing uncertainty and hence more the error.

In this research also, sensitivity analysis is performed because the drivers used in the modelling process are prepared at different resolutions. Therefore to validate the robustness of the model, the model is tried with different combinations of set of drivers. After performing the sensitivity analysis, accuracy assessment of the model is done on the basis of the model accuracy.

10. Accuracy assessment:

The predicted land cover map of the year 1985 along with the raster map of the year 1985 of IGBP dataset is used for calculating the accuracy assessment. The quality of the predicted map is a measure of how well did the model learned from the labelled dataset and how well did it performed on the unlabelled data.

A confusion matrix is created for carrying out the accuracy assessment on the predicted maps. Actual raster map of the year 1985 from IGBP dataset and the predicted land cover map of the year 1985 from the deep neural network is used for generating the confusion matrix. The predicted pixels of the predicted map are then correlated with the actual raster map. The outcome of the accuracy assessment is used for defining the accuracy of the individual classes as well as the overall accuracy. The mathematical equation used for computing the overall accuracy:

$$\boldsymbol{Overall\ accuracy} = \frac{\boldsymbol{Absolute\ number\ of\ the\ correct\ predicted\ samples}}{\boldsymbol{Total\ number\ of\ the\ samples\ in\ the\ map}} * \mathbf{100}\ \%$$

The prediction accuracy of the individual classes is also computed. For computing the accuracy of the individual classes, producer's accuracy and user accuracy is used. Producer's accuracy is computed by considering the correct number of pixels of a particular class divided by the total number of the pixels in the image. It is used for analysing the prediction quality of the deep learning model. It also computes the error of omission. The error of omission is defined as those pixels on the predicted map that are not predicted. Considering those pixels that are predicted correctly for the individual classes and dividing those correct pixels by the total number of predicted pixels of that class defines the user's accuracy. User's accuracy calculates the commission error. It is defined as the probability of the particular pixel belonging to the specific class that it has been predicted to.

User's accuracy = 100% - error of commission

Producer's accuracy = 100% - error of omission

11. Kappa coefficient is used for measuring the accuracy assessment. Kappa coefficient lies in between 0 and 1. The value of 1 indicates that a particular class in predicted map and in the actual map are same and is predicted accurately while 0 indicates that the classes are different and the predicted result is not correct and has 0% accuracy. Kappa coefficient is based on the confusion matrix. The diagonal elements of the matrix represent that the predicted map and the actual map has similar classes.

5. RESULTS AND DISCUSSION:

5.1. Downscaling the drivers:

The drivers or the factors that are responsible for the land cover change are available for the year 2005 and 1995. The drivers are prepared separately for the two time periods i.e. for the year 2005 and for the year 1995. The drivers are Population density, Soil, Rain, Temperature, Road, SRTM (elevation data). All these drivers are prepared at different resolutions. Some of the drivers are present at very coarser resolution and therefore need to be downscaled.

Drivers:

- **Temperature:** The temperature driver is prepared by Indian Meteorological Department (IMD) at 0.5 degree resolution using shepard interpolation technique. For preparing this dataset daily temperature data of 395 weather stations were considered out of 550. These are the stations which have minimum ten years of data and at least for 300 days a year (A. Srivastava et al., 2008). Again interpolating the temperature dataset will not result in much change, moreover, there is not much variability in the study area. The temperature in the study area ranges from 18 degrees to 26 degrees. Therefore the temperature driver is used as it is.

- **Rain:** The rain driver is also prepared by the Indian Meteorological Department (IMD) at 0.25 degree resolution using shepard interpolation technique. For preparing this dataset, daily rainfall data was taken from 6995 rain gauge stations that are located all over India (Pai et al., 2014). Repeatedly interpolating the result will not result in better results. Therefore the rain driver is used as it is.

- **Population:** The population map needs to down scaled. The population in the population map is distributed taluk (district) wise. It means that whole of the taluk has a constant population, while this is not the case in reality, in general, the maximum population is concentrated around the built-ups and as the distance from the built-ups increases the population tends to decrease and is distributed in the whole taluk. Therefore a method is developed to downscale the population map so that maximum population is around the built-ups.

  Methodology for downscaling the population map.

  - All the built-up polygons are extracted from the land cover map.
  - Using the intersect function of ArcGIS 10.1 built-up polygons are made to intersect with the taluk population map.

- Area of all the polygons is computed that lie inside the taluk polygon. Consider this area to be A1. (A1=a1+a2)
- Using multipart to singlepart functionality of ArcGIS10.1, polygons are separated inside the taluk area.
- Individual area of polygons is computed that lie inside the taluk area. For example, two built-up polygons lie inside the taluk region. Area of individual polygon inside the taluk is considered to be a1 and a2.
- For distributing the population inside the taluk following mathematical equation is used:

$$result = \left(\frac{a1}{a1 + a2}\right) * population\ of\ taluk$$

For example: Assume the population of the taluk to be 80. Two built-up polygon lie inside the taluk region. Consider the area of the first polygon to be a1 = 5 and the area of the second polygon to be a2 = 15. According to the above mathematical equation, the distribution of the population for the first built-up polygon is:

$$result = \left(\frac{5}{5+15}\right) * 80 = 20$$

For the second built-up polygon:

$$result = \left(\frac{15}{5+15}\right) * 80 = 60$$

- Using the results computed from the above mathematical equation, these features are then converted into points. A kernel density downscaled population map is prepared from those points. In kernel density functionality of ArcGIS10.1, the values associated with each point are spread from the point location to the specified radius. The density is maximum at the point location and as the distance increases, the value decreases to zero at specified radius. This results in a smooth curved surface over each point (Läuter, 1988). Figure 5.1 shows the downscaled population map using kernel density.

The dark red regions in figure 5.1 clearly indicate the areas of high population density and as the distance from these areas increases the population density tends to decrease. This population map is given as an input to the deep neural network.

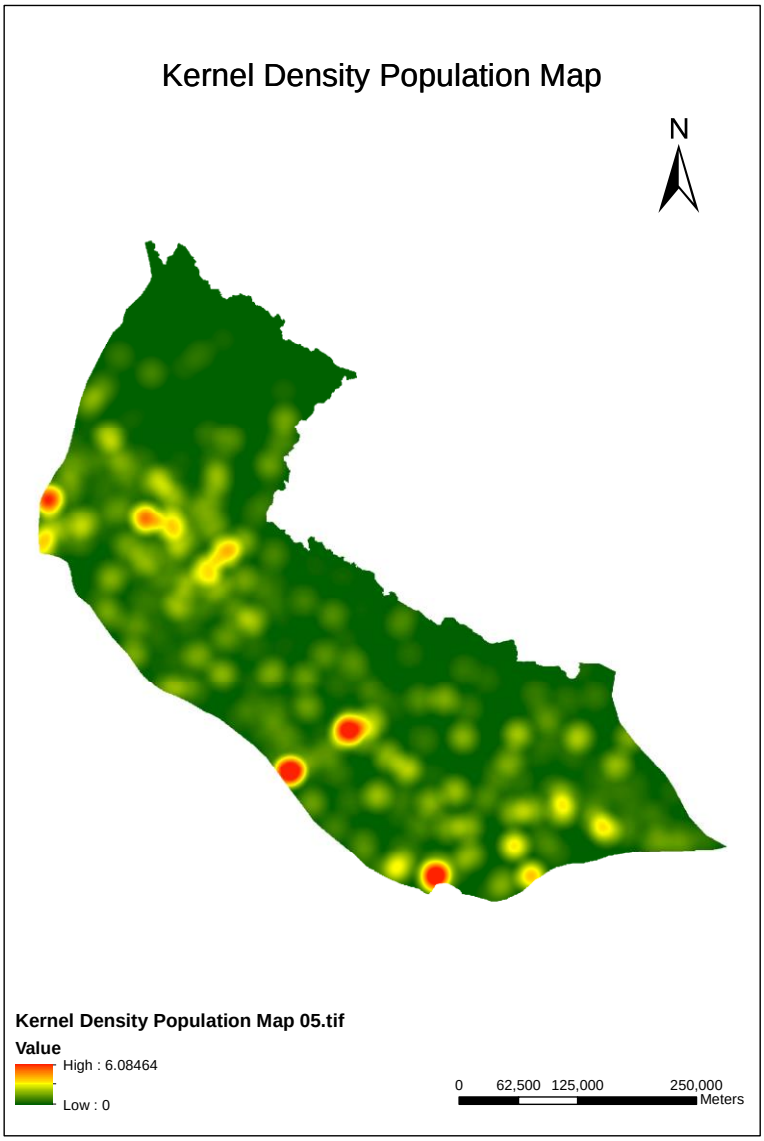


Figure 5.1: Kernel density population map

- **Road**: The roads are considered crucial for the overall development of any area. Settlements or built-up areas grows along the road and its periphery. Road density map is generated from the road map at a cell grid size of 75m using line density function of spatial analyst tool in ArcGIS version 10.1. Line density computes the density of the linear features in the vicinity of each raster cell (Läuter, 1988). Road density map is prepared by using the line density is shown in figure 5.2. The red areas show in figure 5.2 indicates the areas of high road density i.e. it illustrates that in some of the areas, the number of roads intersecting at a point is higher as compared to other areas. This road density map is given as an input to the deep neural network.

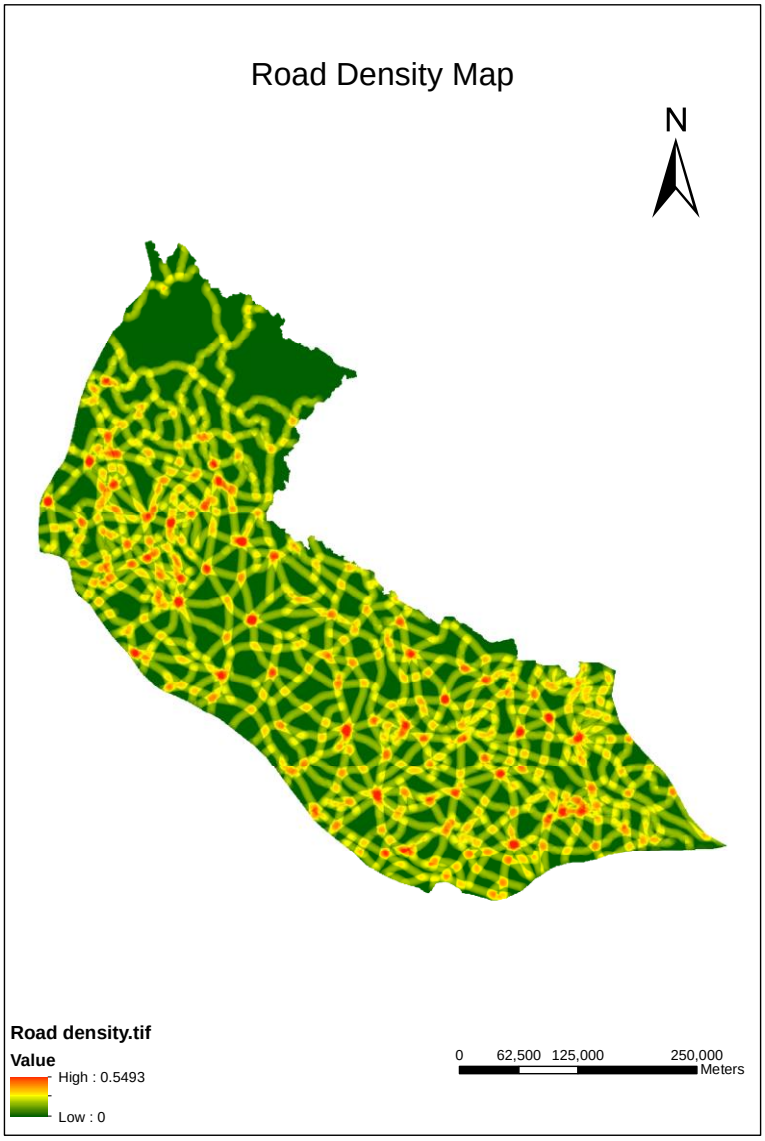


Figure 5.2: Road density map

The relationship between built-up areas and the road map.

The roads are considered vital for the overall development of any area. A relationship is established between the high road density areas and the built-up areas.

The following methodology is adopted for establishing the relationship between the two.

- First, all the built-ups polygons are extracted from the land cover map and saved in a separate shape file
- The built-up shape file is overlaid with the road density map.
- By overlaying the two maps, the relationship between the two is visually analysed and it is observed that, built-ups and roads are directly proportional to each other
- Therefore it is concluded that roads play a vital role in establishing the built-up colonies.

The following maps show the association between the road map and the built-ups.

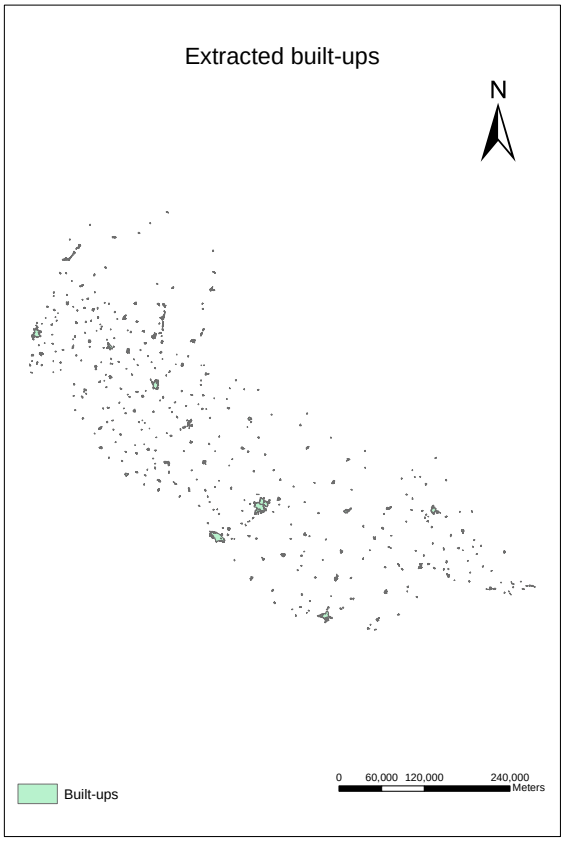


Figure 5.4: Extracted built-ups

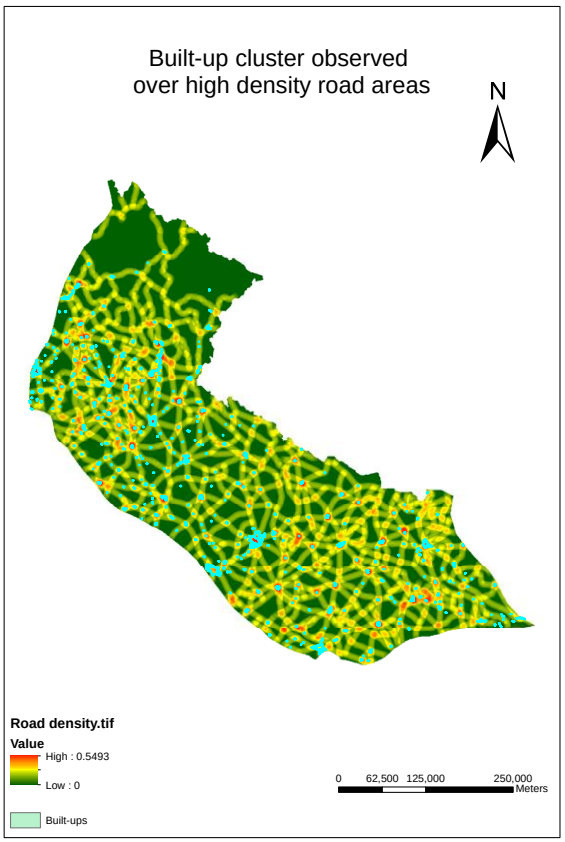


Figure 5.3: Overlay built-up with road density map

Figure 5.4 shows the extracted built-up polygons form the land cover map. In figure 5.3, when all the built-up polygons are selected, shown in blue colour and overlaid with road density map, it is observed that the places where the road density is high built-up clusters are observed. Therefore figure5.3, confirms the relationship between the two.

5.2. Sensitivity Analysis:

The model prepared is tested for different combinations of driver datasets. The sensitivity analysis is performed for the following four cases. The model is trained for the following cases of sensitivity analysis using both the approaches as explained in methodology i.e. using sample data and using complete maps that are given as input to the deep neural network.

S.NO	Cases	Combination of drivers
1	Case 1	Population map, road density map, SRTM map, soil map, rain map, temperature map

2	**Case 2**	Population kernel map, road density map, SRTM map, soil map, rain map, temperature map,
3	**Case 3**	Population kernel map, road density map, SRTM map, soil map
4	**Case 4**	Population kernel map, road density map, SRTM map

5.3. Model parameters:

The deep neural network model is tested for different combinations of the model parameters. The parameters that are varied are as follows:

- Number of iterations
- Number of hidden layers
- Number of neurons in each hidden layer

By varying the above mentioned parameters the loss and accuracy of the model changes. The other parameters stated below are kept same in all the cases.

- Optimiser: The model uses as Adagrad optimiser
- Activation function: Rectified linear unit is used as an activation function.
- Drop out: drop out is used as a regularisation technique.

5.4. Input to deep neural network:

For predicting the land cover map of the year 1985, first the deep neural network is made to learn from the labelled datasets and the drivers i.e. the land cover map of the year 2005 and its corresponding drivers are used as feature columns and the land cover map of the year 1995 is used as target values for the network to learn.

The model learns from the above mentioned datasets for a specified number of iterations. Now the model is tested on the same datasets with which it was made to learn i.e. only the feature columns are provided to the deep neural network and the test accuracy is computed. The test maps are drawn and are compared with actual raster map of the year 1995. If the test accuracy is above a certain threshold, the model is further used for predicting the land cover map of the year 1985.

For predicting the land cover map of the year 1985, the land cover map of the year 1995 and its corresponding drivers are used as feature columns for the network to predict the target classes i.e. the deep neural network predicts the classes of the year 1985 using land cover map of the year 1995 and the driver dataset of the year 1995. The assumption made in predicting the land cover map of the year 1985 is that the drivers of the year 1995 is responsible for the changes in the land cover map of the year 1985.

The following tables represent the different cases of the sensitivity analysis by varying the model parameters.

1) Case 1

1.1. Using sample dataset

S.No	Number of hidden layers	Number of neurons in each hidden layer	Number of iterations	Loss occurred	Test Accuracy
a)	7	7, 14, 21, 28, 32, 42, 15	75000	0.87084	0.69491
b)	16	7, 14, 21, 28, 35, 42, 49, 54, 63, 54, 49, 42, 35, 28, 21, 15	75000	0.27683	0.89285
c)	19	7, 14, 21, 28, 35, 42, 49, 56, 63, 70,77, 63, 54, 49, 42, 35, 28, 21, 15	325000	0.52403	0.78206
d)	21	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 70, 63, 56, 49, 42, 35, 28, 21, 15	150000	0.52897	0.79360
e)	28	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 98, 91, 84, 77, 70, 63, 56, 49, 42, 35, 28, 21, 15	100000	0.29843	0.87958
f)	32	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 112, 119, 112, 105, 98, 91, 84, 77, 70, 63, 56, 49, 42, 35, 28 , 21, 15	150000	0.57547	0.85740
g)	35	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 112, 119, 126, 133, 119, 112, 105, 98, 91, 84, 77, 70, 63, 54, 49, 42, 35, 28, 21, 15	100000	1.15212	0.60820

Table 5-1: Case 1 model parameters using sample dataset

As observed from the above table, initially the number of hidden layers is taken to be seven, since the input to the deep neural network consists of seven maps that are to be learned by the network corresponding to the labelled target data.

When the model is run with seven hidden layers the test accuracy is very low as compared to the loss occurred. As the number of hidden layers is increased from seven to sixteen, the test accuracy of the model is considerably improved while the loss has remarkably reduced for the same number of iterations as compared to the initial seven hidden layer.

As observed from the above table that, the model is tested up to thirty five hidden layers. As the number of hidden layers changes, the test accuracy, and the loss occurred changes for specified number of iterations. The results of all the above mentioned cases i.e. the model accuracy and the loss occurred during training, is considered when there was no longer improvement in model accuracy

For all of the above cases, test maps are prepared corresponding to a different number of hidden layers to visualise, how the neural network is able to retrieve the classes when trained on a sample dataset. Please refer to appendix for all the test maps related to table5-1

When the test maps are observed and visualised, it is found that in all the maps, although the model is able to retrieve all the classes but it is not able to differentiate between the classes, and there is a mismatch among these the classes. The reasons for this mismatch are as follows:

- The model is trained on sample file i.e. the lesser number of pixels are used for training the model. But in reality, the distribution of the classes is biased i.e. is not uniform. Therefore the model is not able to retrieve the classes accurately.

- The drivers, such as temperature and rain are at coarser resolutions as compared to other drivers. But the prediction is carried at much finer resolution.

- It is also observed that although the neural network is able to retrieve all the classes, the network is not able to differentiate between the competitive classes such as crop land, fallow land, grass land and scrub land, plantation land, permanent wet land and wet land. Competitive classes are those classes that are dependent on a similar driver or factor. The above mentioned classes are water dependent, and there is no driver available that is able to differentiate among these classes.

- It is also observed that merely increasing the number of the hidden layers does not result in the improvement of the test accuracy. As the sample dataset does not take into account the driver variability within the sample.

1.2 Case1: complete map are given as input to deep neural network

In this scenario, the complete maps are given as input to the deep neural network without performing the sampling operation. No downscaled drivers were used in this approach. When the model was run using the second approach of methodology, the memory error occurs i.e. it is not able to handle the large size of the datasets.

There are several advantages of using the first approach as compared to the second approach with respect to case1. Since the network is made to run only on the sample datasets, the learning is faster. There is no memory error.

2) Case 2:

2.1) using sample dataset

S.No	Number of hidden	Number of neurons in each hidden layer	Number of	Loss occurred	Test Accuracy

	layers		iterations		
a)	7	7, 14, 21, 28, 32, 42, 15	125000	1.20628	0.57132
b)	16	7, 14, 21, 28, 35, 42, 49, 54, 63, 54, 49, 42, 35, 28, 21, 15	50000	0.35945	0.86036
c)	19	7, 14, 21, 28, 35, 42, 49, 56, 63, 70,77, 63, 54, 49, 42, 35, 28, 21, 15	50000	0.51525	0.80524
d)	21	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 70, 63, 56, 49, 42, 35, 28, 21, 15	100000	1.21618	0.55316
e)	28	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 98, 91, 84, 77, 70, 63, 56, 49, 42, 35, 28, 21, 15	100000	0.87947	0.69996
f)	32	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 112, 119, 112, 105, 98, 91, 84, 77, 70, 63, 56, 49, 42, 35, 28 , 21, 15	50000	0.81811	0.70322
g)	35	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 112, 119, 126, 133, 119, 112, 105, 98, 91, 84, 77, 70, 63, 54, 49, 42, 35, 28, 21, 15	50000	0.98652	0.64541

Table 5-2: Case 2 model parameters using sample dataset

In this case, the model is made to run to using the downscaled population kernel map based on the spatial structure and the road density map, keeping rest of the drivers same as developed by IGBP. Initially, when the network is made to run with seven hidden layers, test accuracy is very poor and the loss occurred is high at one lakh twenty five thousand iterations. Increasing the number of the hidden layer from seven to sixteen it is observed that the test accuracy is remarkably improved and the loss has also reduced. From the test map, it is inferenced that the upper north mountainous region of the study area is learned appreciably. Some of the built-up are also acknowledged.

The model is tested up to thirty five hidden layers for a specified number of iterations. As the number of hidden layers is changed the test accuracy and the loss occurred changes for specified number of iterations. The results of all the above mentioned cases i.e. the model accuracy and the loss occurred during training, is considered when there was no longer improvement in model accuracy.

For all of the above cases, a test map is prepared corresponding to the test accuracy to visualise, how the neural network is able to retrieve the classes when trained on a sample dataset. Please refer to appendix for all the test maps relating to table5-2

When the test maps are observed and visualised, it is found that:

- By downscaling the population map, the network is able to learn the built-up class and establish its relation with the other classes.
- Spatial association of classes is preserved in cases where the testing accuracy is greater than 80%.

- The coarser resolution drivers i.e. the rain and the temperature driver shows the dominance over the other classes.

2.2) Case2: complete map are given as input to deep neural network

Number of hidden layers	7
Number of neuron in each hidden layer	7, 14, 21, 28, 35, 42, 15
Number of iterations	2500
Loss occurred	0.473898
Test accuracy	0.856730

Table5-3: Case 2 model parameters using complete dataset

In the second approach of methodology, all the maps are given as input to the deep neural network. No sampling is performed on these datasets. It is observed that, when the complete maps are given as input to the neural network, the network takes more time to learn as compared to when sampling is performed. The test accuracy achieved by the model is greater than the loss occurred

Test map is generated in order to visualise the test accuracy that the model has achieved. It is observed from the map, that spatial association of the classes is preserved since the network is made to learn from the complete datasets. Although the model is able to preserve the spatial association between the classes but it is not able to reproduce all the classes. Out of fifteen classes, the deep neural network is able to retrieve only eight classes and classes such as built-up, fallow land, plantation land, evergreen broadleaf forest, mixed forest, grass land, permanent wet land have not shown up and these classes are converted into one or the other classes. Such as fallow land is converted into crop land and scrub land, plantation land is converted into crop land and deciduous broadleaf forest, evergreen broadleaf forest is converted into deciduous broadleaf forest, grass land is converted into crop land and permanent wet land is converted into wet land. Please refer to appendix for the test map related to table5-3

In this case, the spatial association of the classes is preserved because the network is able to learn all the possible distributions of the classes within the study area. The reason for some of the classes getting merged into other classes is that the network is not able to establish the relationship between the coarser resolution drivers i.e. the rain and the temperature driver with the other classes. Therefore the spatial variability is not captured.

Further on increasing the number of hidden layers, the python results in memory error i.e. it is not able to accommodate the large size of the matrixes. Therefore with full datasets, the model is not able to handle more than seven hidden layers.

3) Case 3:

3.1) using sample dataset

S.No	Number of hidden layers	Number of neurons in each hidden layer	Number of iterations	Loss occurred	Test accuracy
a)	7	7, 14, 21, 28, 32, 42, 15	100000	0.89059	0.66748
b)	16	7, 14, 21, 28, 35, 42, 49, 54, 63, 54, 49, 42, 35, 28, 21, 15	175000	1.08755	0.58679
c)	19	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 63, 54, 49, 42, 35, 28, 21, 15	75000	0.51549	0.79270
d)	21	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 70, 63, 56, 49, 42, 35, 28, 21, 15	125000	0.59965	0.77283
e)	28	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 98, 91, 84, 77, 70, 63, 56, 49, 42, 35, 28, 21, 15	50000	1.06723	0.64002
f)	32	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 112, 119, 112, 105, 98, 91, 84, 77, 70, 63, 56, 49, 42, 35, 28 , 21, 15	75000	0.81104	0.68197
g)	35	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 112, 119, 126, 133, 119, 112, 105, 98, 91, 84, 77, 70, 63, 54, 49, 42, 35, 28, 21, 15	125000	0.254335	0.89325

Table 5-4: Case 3 model parameters using sample dataset

The model is made to run with downscaled drivers, eliminating the coarser resolution drivers such as rain and the temperature. It is observed that when the rain and the temperature driver are eliminated, the neural network is able to retrieve the classes. The spherical shape around the built-ups is because of the downscale population map that was given as input to the deep neural network. But because of this downscaled population map, the network is able to fetch the built-ups. On increasing the number of hidden layers from seven to sixteen, the loss shoots to high value and the test accuracy reduces. Since in some of the cases such as in case (b) and (e) of table5-4, the loss occurred is high the network is not able to retrieve the classes. Further increasing the hidden layers up to thirty two results in a high loss which is easily visualised in the map. When the model is run with thirty five hidden layers the test accuracy is remarkably improved and the loss is also dropped. The neural network is able to retrieve the classes. The upper northern portion of the study area is appreciably anticipated by the neural network. Please refer to appendix for the test map related to table5-4

From the table 5-4 relating to case 3 of sensitivity analysis, following are the observations:

- The neural network is able to retrieve the classes as compared to case 2 such as built-ups and rivers.
- The spatial association of the classes is not maintained since the network is trained on the sample datasets but in actual the distribution of the classes is biased.

- The reason for the high losses occurred is because the neural network is not able to establish the relationship between the combination of the drivers chosen and the various classes.

3.2) Case3: complete map are given as input to deep neural network

Number of hidden layers	7
Number of neuron in each hidden layer	7, 14, 21, 28, 35, 42, 15
Number of iterations	2500
Loss occurred	0.32689
Test accuracy	0.894420

Table5-5: Case 3 model parameters using complete dataset

In this approach, the complete datasets without sampling are given as input to the deep neural network. The deep neural network takes longer time to learn from the datasets. Please refer to appendix for the test map related to table5-5

It is observed from the test map, when full maps are given as input to the deep neural network. Out of all the fifteen classes the neural network is able to retrieve only ten classes and classes such as built-up class is converted to crop land, evergreen broad leaf forest is converted to evergreen needle forest, mixed forest is converted to scrub land and evergreen needle forest, grass land and permanent wet land are converted into water body. While the pixels of some of the classes got mixed with the other class. For example, some pixels of the fallow land are converted into crop land and evergreen needle forest.

Spatial association of the classes in preserved in this case. The neural network is not able to retrieve all the classes because of the unavailability of the driver to differentiate between the water dependent classes such as permanent wet land and grass land

4) Case 4

4.1) using sample dataset

S.No	Number of hidden layers	Number of neurons in each hidden layer	Number of iterations	Loss occurred	Test accuracy
a)	7	7, 14, 21, 28, 32, 42, 15	100000	1.09786	0.59865
b)	16	7, 14, 21, 28, 35, 42, 49, 54, 63, 54, 49, 42, 35, 28, 21, 15	125000	0.32015	0.86394
c)	19	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 63, 54, 49, 42, 35, 28, 21, 15	100000	0.29376	0.89678
d)	21	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 70, 63, 56, 49, 42, 35, 28, 21, 15	100000	0.39246	0.84961

e)	28	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 98, 91, 84, 77, 70, 63, 56, 49, 42, 35, 28, 21, 15	100000	1.58971	0.41019
f)	32	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 112, 119, 112, 105, 98, 91, 84, 77, 70, 63, 56, 49, 42, 35, 28 , 21, 15	75000	0.46800	0.83284
g)	35	7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91, 98, 105, 112, 119, 126, 133, 119, 112, 105, 98, 91, 84, 77, 70, 63, 54, 49, 42, 35, 28, 21, 15	150000	0.28360	0.88270

Table5-6: Case 4 model parameters using sample dataset

In this case, the model is tested only with continuous data and no categorical data was used for training the network. Initially, when the number of hidden layers is considered to be seven, the loss is very high as compared to the test accuracy. On increasing the number of hidden layers to sixteen, the loss of the model drops and further increasing the hidden layers to nineteen the loss reduces and the test accuracy improves.

It is observed that whenever the loss is very high such as in case (a) and (e), the network is not able to retrieve the classes and the spatial association between the classes is not established. Please refer to appendix for the test map related to table5-6.

From the table 5-6 relating to case 4, following are the observations:

- The neural network is able to retrieve all the classes but it is not able to differentiate between the competitive classes such as fallow land and crop land.
- Neural network is not able to establish the spatial association between the classes, if the loss occurred is very high such as in case (a) and (e) of table5-6

4.2) complete map are given as input to deep neural network

Number of hidden layers	7
Number of neuron in each hidden layer	7, 14, 21, 28, 35, 42, 15
Number of iterations	2500
Loss occurred	0.262978
Test accuracy	0.914166

Table5-7: Case 4 model parameters using complete dataset

From the above table, it is observed, that when complete continuous maps i.e. no categorical driver is given as an input to the neural network, the loss occurred is the lowest of all the cases and the test

accuracy achieved is the highest among all the cases of sensitivity analysis. Please refer to appendix for the test map related to table5-7

It is observed from the test map, that neural network performs best when trained only with the continuous data. Classes such as built-up, crop land, fallow land, plantation land, evergreen needle forest, deciduous broadleaf forest, scrub land, barren land, wet land, waterbody, snow and ice are easily retrieved by the neural network. While the other classes such as evergreen broadleaf forest is converted to evergreen needle forest, mixed forest is converted into scrub land, grass land is converted into water body and barren land, permanent wet land is converted into water body and scrub land.

From the table 5-7, relating to case 4, following are the observations:

- Deep neural network is able to establish the spatial association between the classes i.e. spatial structure of the classes is preserved
- The classes that got converted into other classes is because, in reality, the pixels of these classes are less in number as compared to the other classes. Therefore the neural network is not able to establish the relationship firmly between the classes whose pixels are less in number and the drivers. Moreover, the neural network is not able to distinguish between the competitive classes because of the unavailability of the driver such as scrub land and permanent wet land.

5.5. Summary of the Sensitivity Analysis

Cases	First approach	Second approach
Case 1	Although the neural network is able to retrieve all the classes but it is not able to preserve the spatial structure of the classes because the model is trained on the sample datasets but in reality the distribution of the classes is biased. Learning of neural network is faster,if the model is trained on sample datasets	Python is not able to handle the large size of the datasets, therefore, it results the memory error.
Case 2	Although the neural network is able to retrieve all the classes but it is not able to preserve the spatial structure of the classes because the model is trained on the sample datasets but in reality the distribution of the classes is	Spatial structure of the classes is preserved because the network is able to learn all the possible distributions of the classes. The model is able to retrieve eight classes appreciably out of fifteen classes. Some of the classes that model was not able to retrieve is because of the unavailability of the driver to distinguish between the competitive

	biased. Learning of neural network is faster if the model is trained on sample datasets	classes and the water dependent classes and also because of the lesser number of pixel count of some of the classes such as evergreen broadleaf forest. The model is tested for seven hidden layers only. Test accuracy achieved with seven hidden layer is 85.6%. If the model is tested for more than seven hidden layers, the python results in memory error.
Case 3	Although the neural network is able to retrieve all the classes but it is not able to preserve the spatial structure of the classes because the model is trained on the sample datasets but in reality the distribution of the classes is biased. Learning of neural network is faster if the model is trained on sample datasets	Spatial structure of the classes is preserved because the network is able to learn all the possible distributions of the classes. The model is able to retrieve ten classes appreciably out of fifteen classes. Some of the classes that model was not able to retrieve is because of the unavailability of the driver to distinguish between the competitive classes and also because of the lesser number of pixel count of some of the classes such as evergreen broadleaf forest. The model is tested for seven hidden layers only. Test accuracy achieved is 89.4%. If the model is tested for more than seven hidden layers, the python results in memory error.
Case 4	Although the neural network is able to retrieve all the classes but it is not able to preserve the spatial structure of the classes because the model is trained on the sample datasets but in reality the distribution of the classes is biased. Learning of neural network is faster, if the model is trained on sample datasets	Spatial structure of the classes is preserved because the network is able to learn all the possible distributions of the classes. The model is able to retrieve eleven classes appreciably out of fifteen classes. Some of the classes that model was not able to retrieve is because of the unavailability of the driver to distinguish between the water dependent classes such as permanent wet land and also because of the lowest number of the pixel count of some of the classes such as evergreen broadleaf forest. The model is tested for seven hidden layers only. Test accuracy achieved is 91.4% If the model is tested for more than seven hidden layers, the python results in memory error.

Table5-8: Summary of sensitivity analysis

From the above formulated table 5-8, it is observed that the model performed best in case 4 using the second approach of methodology i.e. when complete maps are given as input to the deep neural network. The model achieved test accuracy of 91.4%. Therefore this model is used for predicting the land cover map of the year 1985.

5.6. Prediction:

For predicting the land cover map of the year 1985, the model developed in **case 4** by training the deep neural network on complete datasets is used. The deep neural network is fed with the land cover map of the year 1995 and its corresponding driver. It is assumed that the drivers of the year 1995 are responsible for the changes in the land cover map of the year 1985. Based on the previous learning of the model on different datasets i.e. earlier the model was trained on the land cover map of the year 2005 and 1995 and the driver dataset of the year 2005, the neural network is used for predicting the land cover map of the year 1985 using land cover map of the year 1995 and its drivers.

Figure 5.6 is used for visualising the predicted land cover map of the year 1995 generated by the model and compare it with the original land cover map of the year 1985.

From the predicted land cover map of the year 1985, following are the observations:

- Out of the fifteen different land cover classes, the model is able to predict eleven classes appreciably.

- It is observed that built-ups are less in number in the predicted land cover map as compared to original land cover map of the year 1985. This is because, in reality, the built-ups are less in number in the original raster map of the year 1985 as compared to the land cover map of the year 1995 and 2005. Moreover, it is assumed that the road driver which is available for the year 2005 only, is responsible for the changes in the land cover map of the year 1985 but in reality, this is not the case. Therefore neural network is not able to establish the sturdy spatial association between the roads and the built-ups. But still, the model is able to predict some of the built-ups with which it finds the spatial association.

- The neural network is not able to differentiate between the different forest types. For example, evergreen broadleaf forest is predicted as scrub land and evergreen needle forest. In the actual raster land cover map of the year 1985, evergreen broadleaf forest has the least number of the pixel count among all the classes and most of it is either surrounded by scrub land, evergreen needle forest or barren land. Therefore neural network is not able to predict the evergreen broadleaf forest correctly. This is because the association of the classes is dependent on the number of pixels of a particular class that the model is able to learn. Class of mixed forest is predicted as scrub land. It is observed that most of the mixed forest is surrounded by the scrub land. Therefore the model has predicted mixed forest as scrub land

- Class of grass land is predicted as water body. Grass land act as a competitive class. Therefore neural network is not able to predict the grass land correctly.

- Class of permanent wet land is predicted as water body because of the unavailability of the driver to differentiate between the different water dependent classes.

- Rest of the classes are predicted appreciably

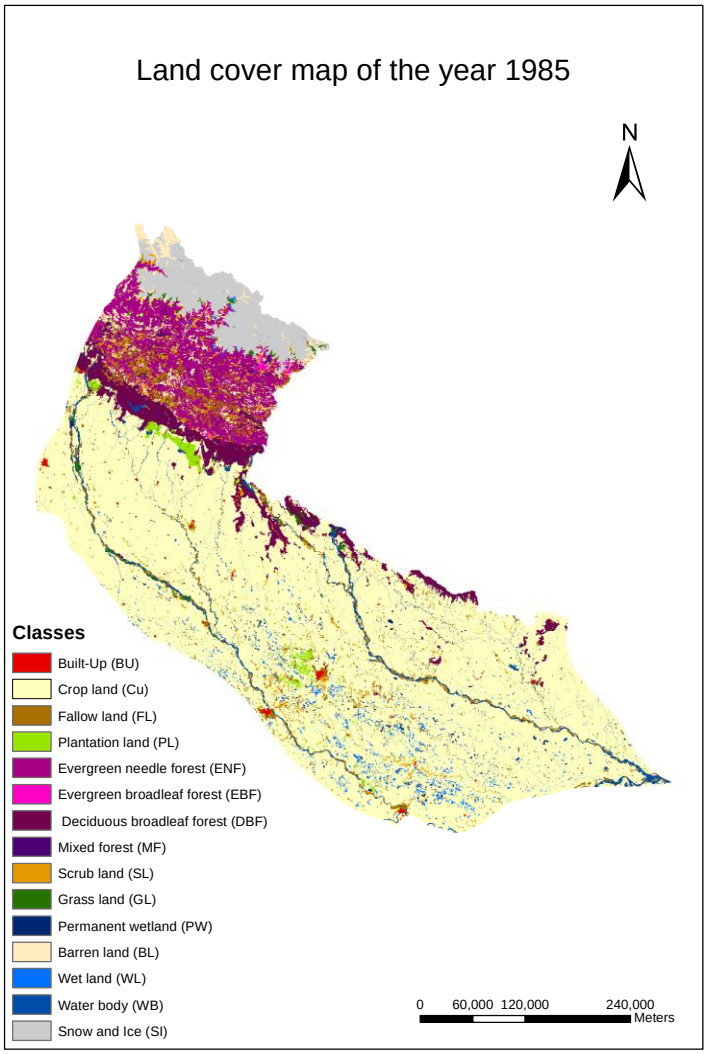


Figure5.5: Actual raster land cover map of year 1985

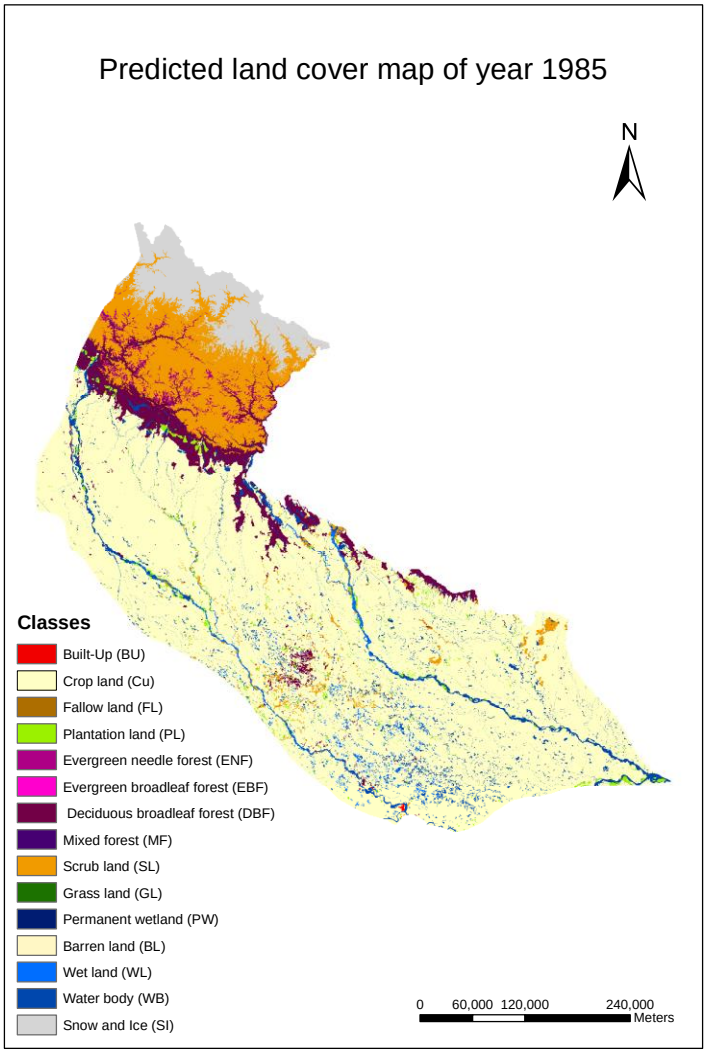


Figure5.6: Predicted land cover map of year 1985

5.7. Accuracy Assesment:

To access the accuracy of the predicted image, the confusion matrix is created. In confusion matrix, the predicted land cover map of the year 1985 is compared with available raster map of the year 1985. Table 5.9 shows the confusion matrix.

Classes	BU	CL	FL	PL	ENF	EBF	DBF	MF	SL	GL	PW	BL	WL	WB	SI
BU	2830	147290	538	0	3855	0	64	0	706	0	0	0	0	0	0
CL	0	23924418	503	162608	234820	0	49744	0	128384	0	0	888	30503	32078	401
FL	1714	105401	775	234878	162066	0	38410	0	386066	0	0	244	20439	67693	54
PL	538	12913	0	3349	0	0	301884	0	1273	0	0	0	343	1074	0
ENF	0	0	0	0	0	0	166887	0	2193080	0	0	0	0	0	198101
EBF	0	0	0	0	0	0	4611	0	50688	0	0	0	0	0	2252
DBF	0	0	0	0	0	0	1303987	0	246938	0	0	0	0	498	0
MF	0	252	0	0	0	0	7769	0	156451	0	0	0	216	1764	232
SL	0	19847	0	7862	0	0	24453	0	557358	0	0	0	24372	306286	22081
GL	0	8468	0	0	0	0	0	0	44778	0	0	0	5693	86567	14780
PW	0	9212	0	3033	0	0	323	0	4431	0	0	0	361	184139	0
BL	0	0	0	0	0	0	0	0	115101	0	0	0	1	810	225739
WL	0	2162	0	0	0	0	175	0	44045	0	0	0	237902	136908	4560
WB	26	86291	31	33220	0	0	381	0	26308	0	0	0	255472	491631	1269
SI	0	0	0	0	0	0	0	0	161153	0	0	0	0	0	1673375

Table 5-9: Confusion matrix

5.8. User and Producer accuracy

Classes	User's accuracy (%)	Producer accuracy (%)
Built-up	18.2%	55.40%
Crop land	97.39%	98.39%
Fallow land	0.08%	41.96%
Plantation	10.4%	7.5%
Ever green needle forest	0%	0%
Ever green broad leaf forest	0%	NA
Deciduous broad leaf forest	84.05%	68.68%
Mixed forest	0%	NA
Scrub land	57.92%	13.54%
Grass land	0%	NA
Permanent wet land	0%	NA
Barren land	0%	0%
Wet land	55.88%	41.35%
Water body	54.95%	37.54%
Snow and Ice	91.22%	78.09%

Table 5-10: User and producer accuracy

User and producer accuracy are calculated based on the confusion matrix. The model is able to predict the eleven classes appreciably out of fifteen classes. The four classes that model was not able to predict are

evergreen broadleaf forest, mixed forest, grass land and permanent wet land. Therefore these classes are considered as NA (not applicable) and do not contribute in computing the accuracy. But according to the table5-10 the user accuracy of evergreen needle forest and barren land is also 0%. This is because even if the model is able to predict these classes but the location of these classes are changed in the predicted map and therefore the accuracy of both these classes is 0%. Since the distribution of the all the classes is not uniform therefore the less dominant classes gets converted into more dominant classes. For example, evergreen broadleaf forest get converted into evergreen needle forest.

Overall accuracy computed from the confusion matrix is 80.07%
Kappa coefficient calculated for the predicted land cover map of the year 1985 is 60.73%

6. CONCLUSION AND RECOMMENDATIONS

6.1. Conclusion

In this research work, a machine learning technique i.e. deep neural net is explored for predicting the spatial structure to explain the trends or pattern for the land cover change. For prediction, the land cover maps of the year of the year 2005 and 1995 and the driver dataset of the year 2005 and 1995 are available. The model designed, is trained on the above mentioned dataset and if the model accuracy is above a certain threshold, the model is further used for predicting the land cover map of the year 1985.

Land cover change is the result of the various drivers or factors that bring about this change. In this study, drivers such as population, road, SRTM (elevation data), rain, soil, and temperature are considered. Sensitivity analysis is performed to assess the robustness of the model developed for a different combination of the set of drivers. It is found in this research that the model performed best when trained with continuous data i.e. population kernel map, road density, and elevation data.

Two different methods are considered for training the deep neural network. In the first approach, the model is trained on the sample datasets created from the study area and in the second approach the model is trained on the complete datasets i.e. when no sampling is performed.

It is observed that when the model is trained with the continuous datasets using the sample data, the model is able to retrieve all the classes but it is not able to establish the relationship between the drivers and the land cover classes. This is because the model is trained using the sample dataset but in reality, the distribution of the various classes is biased i.e. is not uniform. But when the model is trained with continuous data using complete datasets, the model achieved the test accuracy of 91.4% corresponding to the case 4 of sensitivity analysis. Of all the fifteen classes the model is able to retrieve the eleven classes appreciably and the classes that the model is not able to retrieve is because no driver is available to differentiate between the water dependent classes such as permanent wet land and scrub land. Class of evergreen broadleaf forest was predicted by the model as scrub land and evergreen needle forest because of the least number of the pixels of evergreen broadleaf forest among all the classes and most of it is surrounded by scrub land.

The overall accuracy achieved by the model is 80.7% and the kappa accuracy achieved is 60.73% because the model is able to predict eleven classes appreciably. The pixels of the remaining four classes were predicted as pixels of the other classes such as permanent wet land is predicted as water body. This is due to the unavailability of the driver to differentiate between the competitive classes and it also depends on the pixel count of the particular class that is learned by the model. Apart from the four classes that the

model was not able to predict, the user accuracy of the barren land and evergreen needle forest is computed as 0% because although the model was able to predict the class of barren land and evergreen needle forest but the location of the pixels of these classes are changed in the predicted map.

6.2. Answers to research question

1. **What are the drivers or the factors that affect the land change over time?**

   Answer: The drivers or the factors that are available for carrying out the research are population, road, SRTM, soil, rain and temperature. These drivers in one or the other way effect the land cover change over time, but since these drivers are prepared at different resolutions the significance of each driver is found by sensitivity analysis. By performing the sensitivity analysis, it is found that the model performs best with downscaled population map, SRTM map and road density map. The model achieved the highest test accuracy of 91.4% with the above mentioned set of drivers.

2 **How will the spatial pattern such as areas that are adjacent or nearby to one another, lead to the prediction of spatial structure?**

   Answer. The various classes of the land cover map have a spatial pattern. For example: in the land cover map it is observed that in general, the scrub land is found around the water bodies and in the mountain regions. In the upper northern region of the study area i.e. the mountainous area, most of the forest is surrounded by scrub land i.e. scrub land has a larger number of pixel count than forest areas. Therefore in the predicted map, the deep neural network has predicted most of the forest areas as scrub land. This is because the association of the classes is dependent on the number of pixels of a particular class that the model is able to learn.

3 **What is the relation between spatial structure and land change drivers?**

   Answer: Relationship between land change drivers and spatial structure is found as explained in section 5.1, it was found that built-up areas are observed over the high road density areas. Deep neural network establishes the spatial association between the various classes and land change drivers that aid in predicting the land cover map of year 1985

6.3. Recommendations:

1. Some of the drivers that are used for predicting the land cover map of the year the 1995, are very coarser resolution such as rain and temperature. Generation of the driver dataset at a finer resolution may result in better accuracy of the predicted map.

2. More cases can be formulated of the deep neural network to look forward to any trend or pattern present in the data.
3. All the cases from table number 5-1 to 5.7 are performed on CPU. The GPU version should also be tested for predicting the land cover map of the year 1995. The GPU learns from the datasets faster than CPU.

LIST OF REFERENCES

Alvarez Martinez, J. M., Suarez-Seoane, S., & De Luis Calabuig, E. (2011). Modelling the risk of land cover change from environmental and socio-economic drivers in heterogeneous and changing landscapes: The role of uncertainty. *Landscape and Urban Planning*, *101*(2), 108–119. https://doi.org/10.1016/j.landurbplan.2011.01.009

Andrew Ng, C. S. at B. (2015). Extract Data Conference | SlideShare. Retrieved February 24, 2017, from https://www.slideshare.net/ExtractConf

Arsanjani, J. J., Helbich, M., Kainz, W., & Boloorani, A. D. (2012). Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *International Journal of Applied Earth Observation and Geoinformation*, *21*(1), 265–275. https://doi.org/10.1016/j.jag.2011.12.014

Bharati, L. and Jayakody, P. (2010). Hydrology of the Upper Ganga River. *International Water Management Institute*, 1–24. Retrieved from http://publications.iwmi.org/pdf/H043412.pdf

Brink, A. B., Bodart, C., Brodsky, L., Defourney, P., Ernst, C., Donney, F., … Tuckova, K. (2014). Anthropogenic pressure in East Africa-Monitoring 20 years of land cover changes by means of medium resolution satellite data. *International Journal of Applied Earth Observation and Geoinformation*, *28*(1), 60–69. https://doi.org/10.1016/j.jag.2013.11.006

Burg, M. B. (2016). Uncertainty and Sensitivity Analysis in Archaeological Computational Modeling. https://doi.org/10.1007/978-3-319-27833-9

Chou, Y. (1995). Spatial pattern and spatial autocorrelation. *Spatial Information Theory A Theoretical Basis for GIS*. https://doi.org/10.1007/3-540-60392-1_24

Dewan, A. M., & Yamaguchi, Y. (2009). Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization. *Applied Geography*, *29*(3), 390–401. https://doi.org/10.1016/j.apgeog.2008.12.005

Du, K.-L., & Swamy, M. N. S. (2014). *Neural Networks and Statistical Learning*. https://doi.org/10.1007/978-1-4471-5571-3

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Feedforward Networks (pp. 168–227). Retrieved from http://www.deeplearningbook.org/contents/mlp.html

GoogleResearch. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Retrieved from http://download.tensorflow.org/paper/whitepaper2015.pdf

Gounaridis, D., Apostolou, A., Riga, M., Moustos, P., Bika, K., Kontou, D., & Koukoulas, S. (2016). Change detection of forests and semi-natural areas in Greece for the period 1990-2010. *AEgean Journal of Environmental Sciences*, (January).

Haykin, S. (1999). Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7. *The Knowledge Engineering Review*. https://doi.org/10.1017/S0269888998214044

Kakaraparthi, S. K., & Kockelman, K. M. (2011). An Application of UrbanSim to the Austin, Texas

Region: Integrated-Model forecasts for the year 2030. *Journal of Urban Planning and Development*, *137 (3)*, 238–247. https://doi.org/10.1017/CBO9781107415324.004

Läuter, H. (1988). Silverman, B. W.: Density Estimation for Statistics and Data Analysis. Chapman & Hall, London – New York 1986, 175 pp., £12.—. *Biometrical Journal*, *30*(7), 876–877. https://doi.org/10.1002/bimj.4710300745

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Do Deep Convolutional Nets Really Need to be Deep and Convolutional?, *521*. https://doi.org/10.1038/nature14539

Li, X., & Yeh, A. G. O. C. N.-L.-2002-01 L. B.-L. (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS. *Int, J.Geographical Information Science*, *16*(4), 323–343. https://doi.org/10.1080/13658810210137004

Liu, X., Li, X., Shi, X., Wu, S., & Liu, T. (2008). Simulating complex urban development using kernel-based non-linear cellular automata. *Ecological Modelling*, *211*(1–2), 169–181. https://doi.org/10.1016/j.ecolmodel.2007.08.024

Maithani, S. (2009). RESEARCH ARTICLE A Neural Network based Urban Growth Model of an Indian City. *Journal of the Indian Society of Remote Sensing*, *2021*(September), 363–376.

Newman, P. (2006). The environmental impact of cities. *Environment and Urbanization*, *18*(2Newman, P. (2006). The environmental impact of cities. Environment and Urbanization, 18(2), 275–295. doi:10.1177/0956247806069599), 275–295. https://doi.org/10.1177/0956247806069599

Nogueira, K., Penatti, O. A. B., & Jefersson, A. (2016). Author ' s Accepted Manuscript Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification Reference : To appear in : Pattern Recognition Towards Better Exploiting Convolutional Neural Networks for. *Pattern Recognition*, 1–18. https://doi.org/10.1016/j.patcog.2016.07.001

Pai, D. S., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S., & Mukhopadyay, B. (2014). Development of a new high spatial resolution (0 . 25 ° × 0 . 25 °) Long Period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam*, *65*(1), 1–18.

Parker, D. C., Manson, S. M., Janssen, M. a., Hoffmann, M. J., & Deadman, P. (2002). Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change : A Review Forthcoming. *Workshop on Agent-Based Models of Land Use*, (Nag 56406), 75. https://doi.org/10.1111/1467-8306.9302004

Prafull. (2015). Indian River Systems | IAS. Retrieved February 16, 2017, from https://ias.org.in/2015/10/indian-river-systems.html

Raut, S. (2017). What makes Deep Learning deep....and world-changing? Retrieved May 19, 2017, from http://readwrite.com/2017/04/28/what-is-deep-learning-il1/

Ruder, S. (2016, September 15). An overview of gradient descent optimization algorithms. Retrieved May 30, 2017, from http://arxiv.org/abs/1609.04747

Rujoiu-Mare, M.-R., & Mihai, B.-A. (2016). Mapping Land Cover Using Remote Sensing Data and GIS Techniques: A Case Study of Prahova Subcarpathians. *Procedia Environmental Sciences*, *32*, 244–255.

https://doi.org/10.1016/j.proenv.2016.03.029

Schneider, A. (2012). Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sensing of Environment*, *124*, 689–704. https://doi.org/10.1016/j.rse.2012.06.006

Sohl, T., & Sleeter, B. (2012). Role of Remote Sensing for Land-Use and Land-Cover Change Modeling (pp. 225–240). https://doi.org/10.1201/b11964-18

Srivastava, A., Rajeevan, M., & Kshirsagar, S. R. (2008). *Development of a High Resolution Daily Gridded. NCC Research Report.*

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958. https://doi.org/10.1214/12-AOS1000

TechTarget. (2016). What is machine learning? - Definition from WhatIs.com. Retrieved February 24, 2017, from http://whatis.techtarget.com/definition/machine-learning

Tompson, J., Jain, A., LeCun, Y., & Bregler, C. (2014). Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation. *Advances in Neural Information Processing Systems*, 1799--1807. Retrieved from http://arxiv.org/abs/1406.2984

United Nations. (2014). Department of economic and social affairs of the united nations population division (UNPD) (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352). *New York, United*, 32. https://doi.org/10.4054/DemRes.2005.12.9

Use, L., Cover, L., & Change, L. C. (2015). ISRO-GBP PROJECT Human Dimensions in Indian River Basins ISRO-GBP PROJECT Land Use / Land Cover Dynamics and Impact of, 1–7.

Veldkamp, A., & Fresco, L. O. (1996). CLUE: a conceptual model to study the Conversion of Land Use and its Effects. *Ecological Modelling*, *85*(2–3), 253–270. https://doi.org/10.1016/0304-3800(94)00151-0

Vermeiren, K., Vanmaercke, M., & Van Rompaey, A. (2016). ASSURE: a model for the simulation of urban expansion and intra-urban social segregation. *Internation Journal of Geographical Information Science*, *30*(8), 1–24. https://doi.org/10.1080/13658816.2016.1177641

Yang, X., Zheng, X.-Q., & Lv, L.-N. (2012). A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata. *Ecological Modelling*, *233*, 11–19. https://doi.org/10.1016/j.ecolmodel.2012.03.011

Yu, Y., Li, J., Guan, H., & Wang, C. (2016). Automated Detection of Three-Dimensional Cars in Mobile Laser Scanning Point Clouds Using DBM-Hough-Forests. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(7), 4130–4142. https://doi.org/10.1109/TGRS.2016.2537830

Zessner, M., Schönhart, M., Parajka, J., Trautvetter, H., Mitter, H., Kirchner, M., … Schmid, E. (2016). A novel integrated modelling framework to assess the impacts of climate and socio-economic drivers on land use and water quality. *Science of The Total Environment*, *579*, 1137–1151. https://doi.org/10.1016/j.scitotenv.2016.11.092

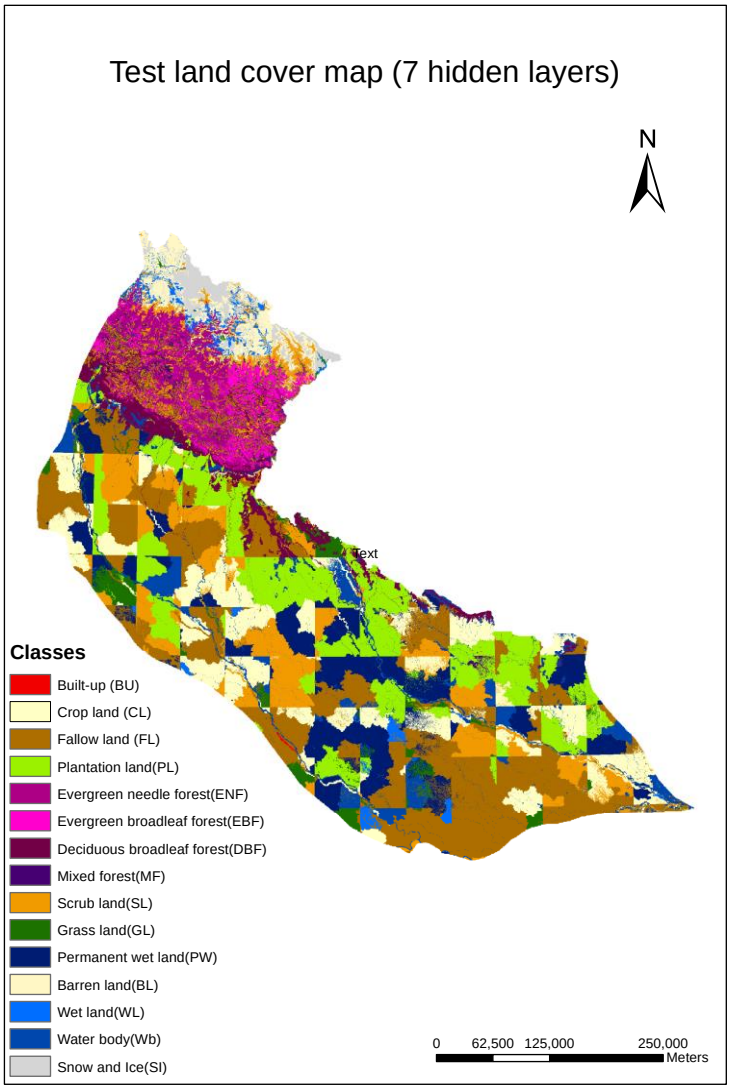
Zhang, L., Xia, G., Wu, T., Lin, L., & Tai, X. C. (2015). Deep Learning for Remote Sensing Image Understanding, *2016*.

Zhang, X., Kang, T., Wang, H., & Sun, Y. (2010). Analysis on spatial structure of landuse change based on remote sensing and geographical information system. *International Journal of Applied Earth Observation and Geoinformation*, *12*(SUPPL. 2), S145–S150. https://doi.org/10.1016/j.jag.2010.04.011

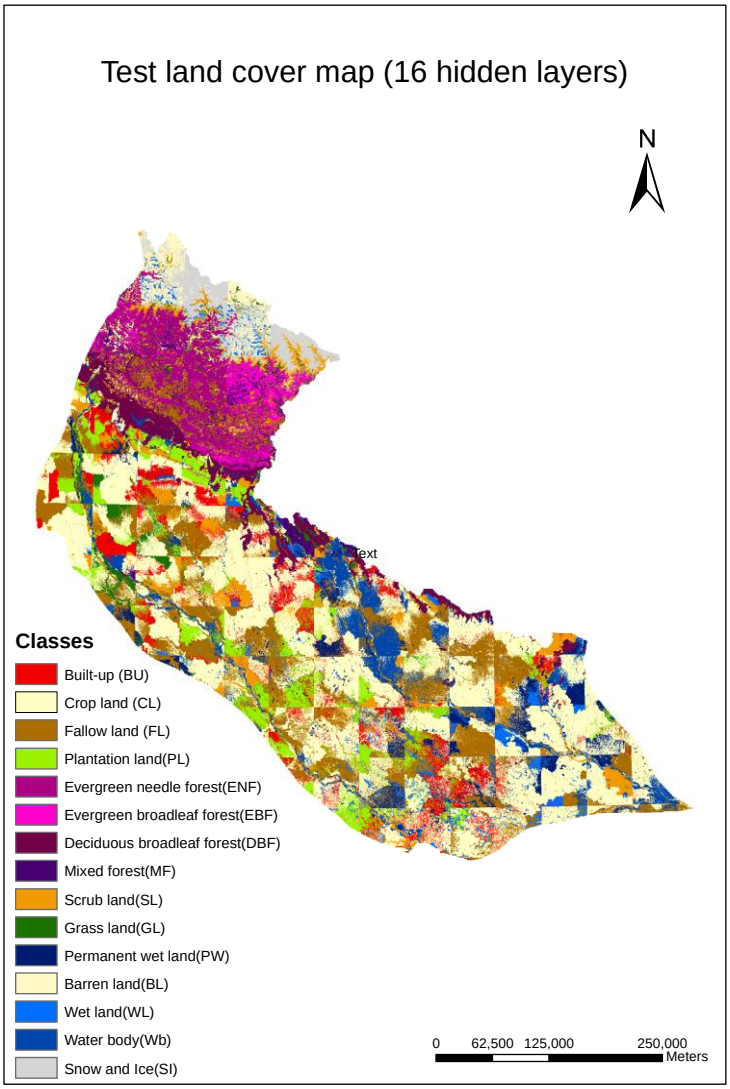
Zhao, W., & Du, S. (2016). Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *113*, 155–165. https://doi.org/10.1016/j.isprsjprs.2016.01.004

7. APPENDIX

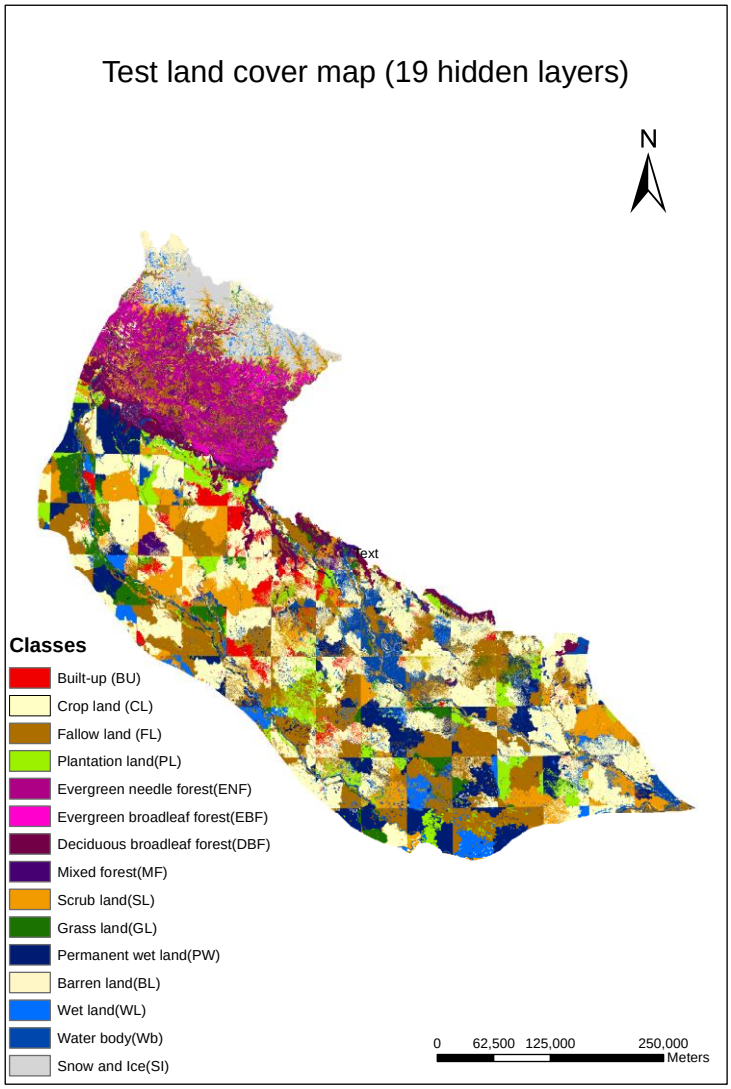
Test Maps of Case 1: table 5-1



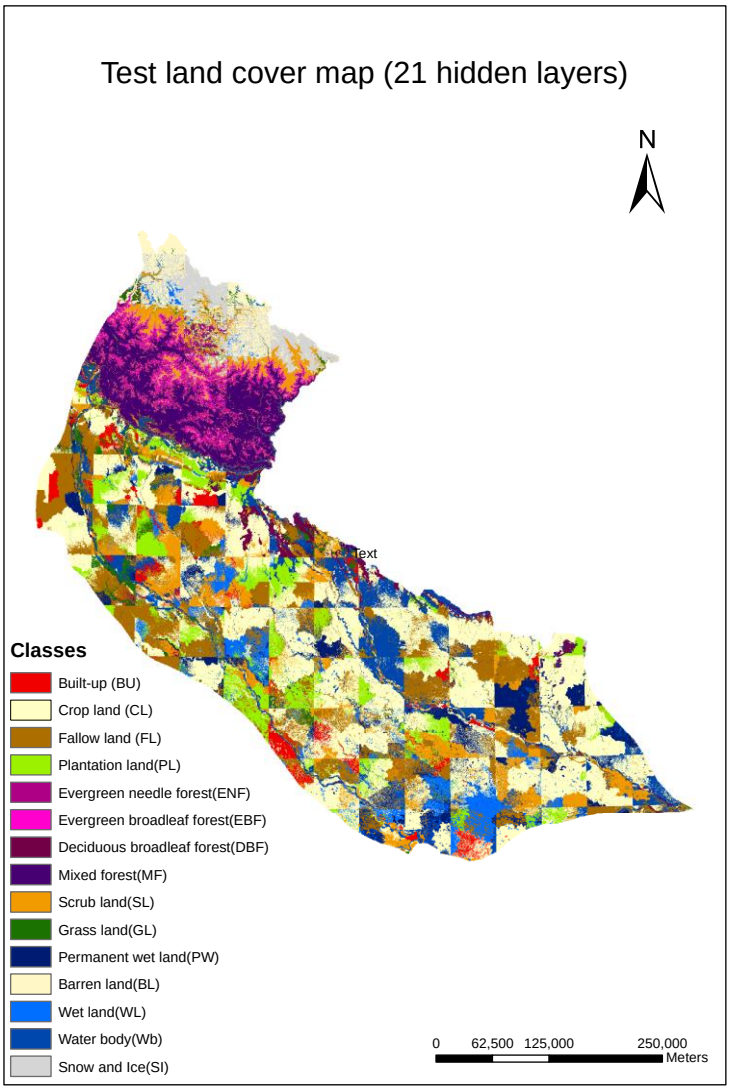
(a)



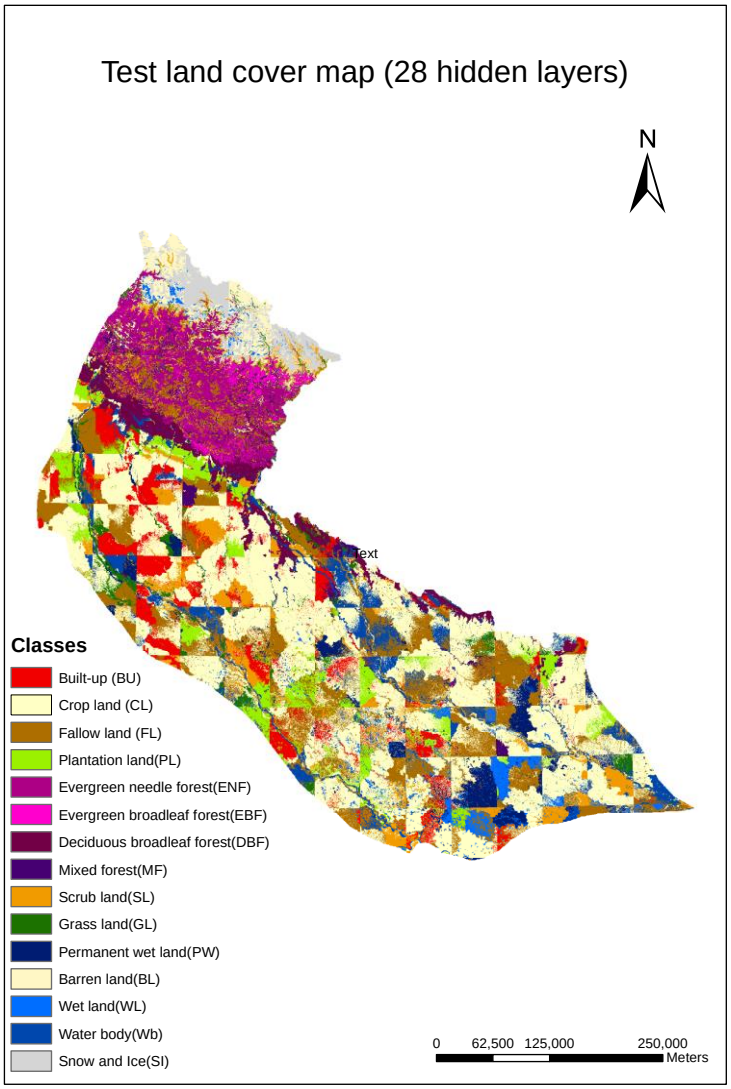
(b)


Test land cover map (19 hidden layers)
N
Text
Classes
Built-up (BU)
Crop land (CL)
Fallow land (FL)
Plantation land(PL)
Evergreen needle forest(ENF)
Evergreen broadleaf forest(EBF)
Deciduous broadleaf forest(DBF)
Mixed forest(MF)
Scrub land(SL)
Grass land(GL)
Permanent wet land(PW)
Barren land(BL)
Wet land(WL)
Water body(Wb)
Snow and Ice(SI)
0 62,500 125,000 250,000
Meters

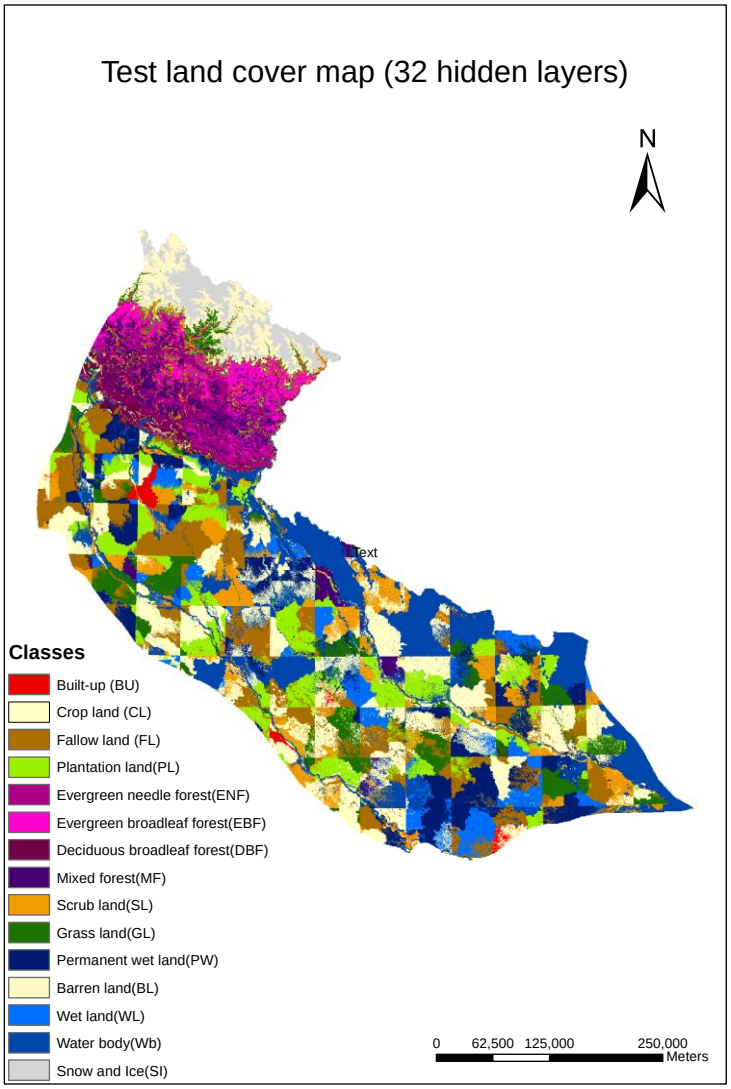
(c)


Test land cover map (21 hidden layers)
N
Text
Classes
Built-up (BU)
Crop land (CL)
Fallow land (FL)
Plantation land(PL)
Evergreen needle forest(ENF)
Evergreen broadleaf forest(EBF)
Deciduous broadleaf forest(DBF)
Mixed forest(MF)
Scrub land(SL)
Grass land(GL)
Permanent wet land(PW)
Barren land(BL)
Wet land(WL)
Water body(Wb)
Snow and Ice(SI)
0 62,500 125,000 250,000
Meters

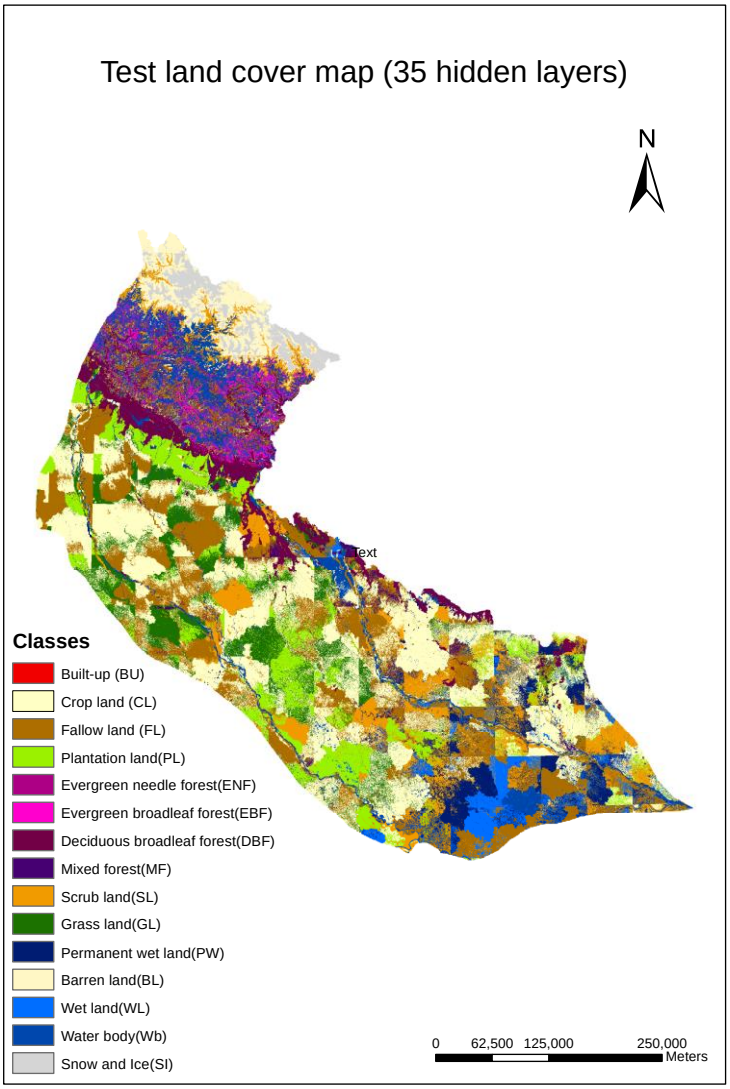
(d)


Test land cover map (28 hidden layers)
N
Text
Classes
Built-up (BU)
Crop land (CL)
Fallow land (FL)
Plantation land(PL)
Evergreen needle forest(ENF)
Evergreen broadleaf forest(EBF)
Deciduous broadleaf forest(DBF)
Mixed forest(MF)
Scrub land(SL)
Grass land(GL)
Permanent wet land(PW)
Barren land(BL)
Wet land(WL)
Water body(Wb)
Snow and Ice(SI)
0 62,500 125,000 250,000
Meters

(e)

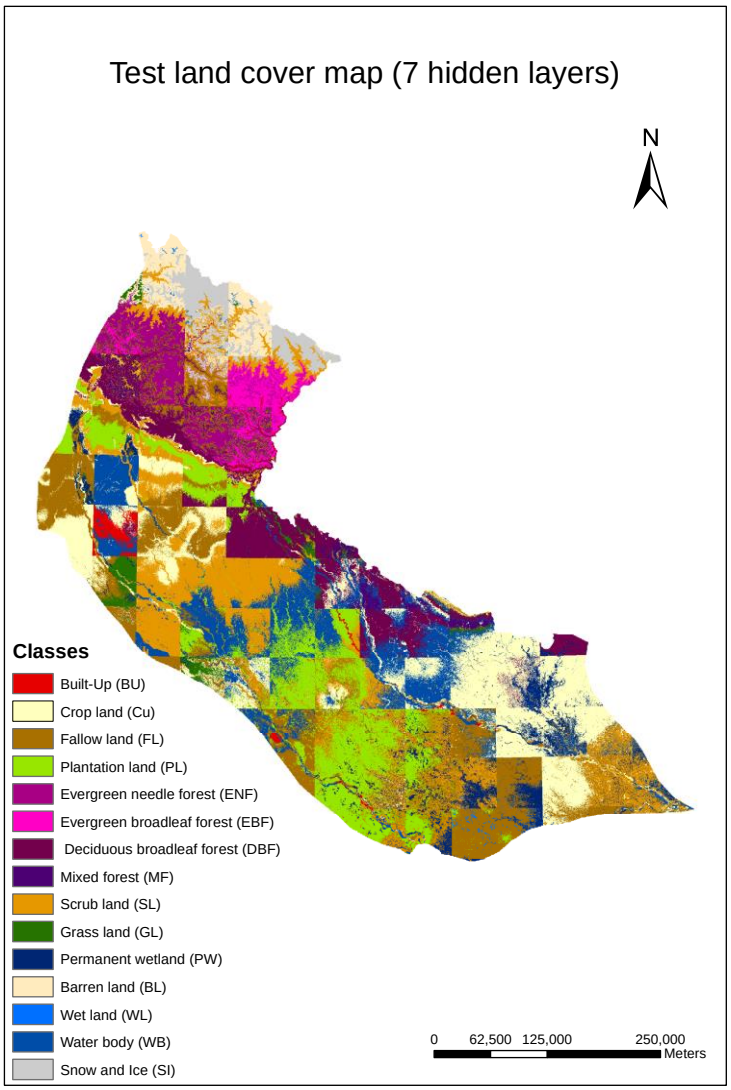

Test land cover map (32 hidden layers)
N
Text
Classes
Built-up (BU)
Crop land (CL)
Fallow land (FL)
Plantation land(PL)
Evergreen needle forest(ENF)
Evergreen broadleaf forest(EBF)
Deciduous broadleaf forest(DBF)
Mixed forest(MF)
Scrub land(SL)
Grass land(GL)
Permanent wet land(PW)
Barren land(BL)
Wet land(WL)
Water body(Wb)
Snow and Ice(SI)
0 62,500 125,000 250,000
Meters

(f)

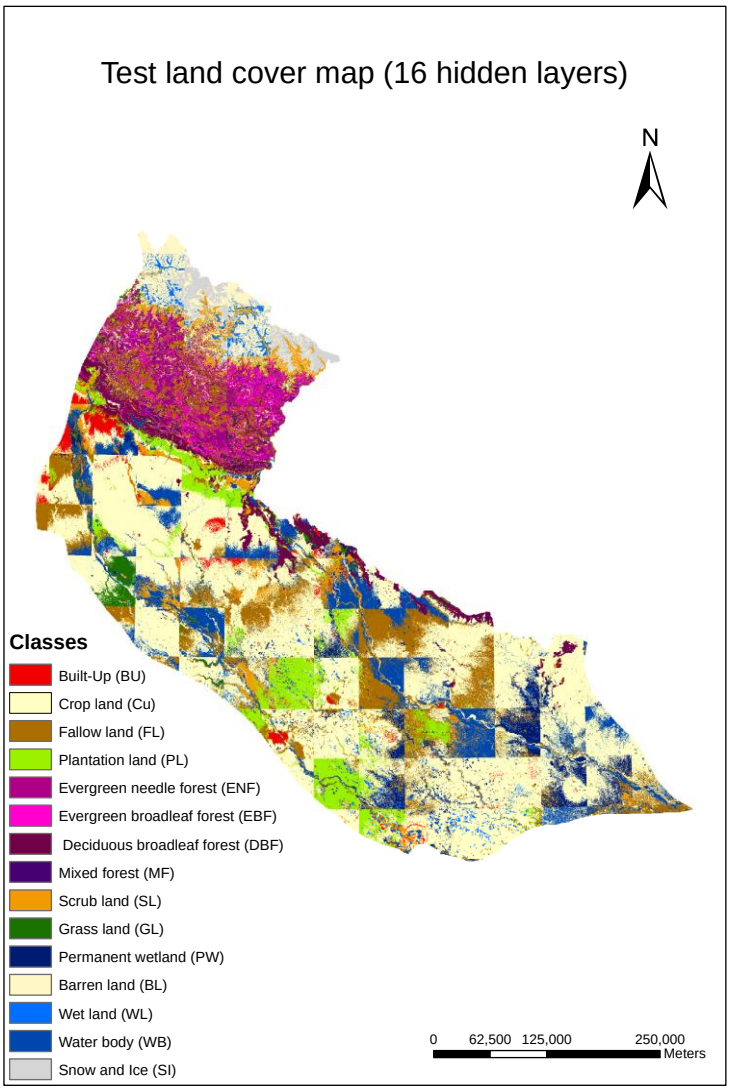

Test land cover map (35 hidden layers)
N
Text
Classes
Built-up (BU)
Crop land (CL)
Fallow land (FL)
Plantation land(PL)
Evergreen needle forest(ENF)
Evergreen broadleaf forest(EBF)
Deciduous broadleaf forest(DBF)
Mixed forest(MF)
Scrub land(SL)
Grass land(GL)
Permanent wet land(PW)
Barren land(BL)
Wet land(WL)
Water body(Wb)
Snow and Ice(SI)
0 62,500 125,000 250,000
Meters

(g)

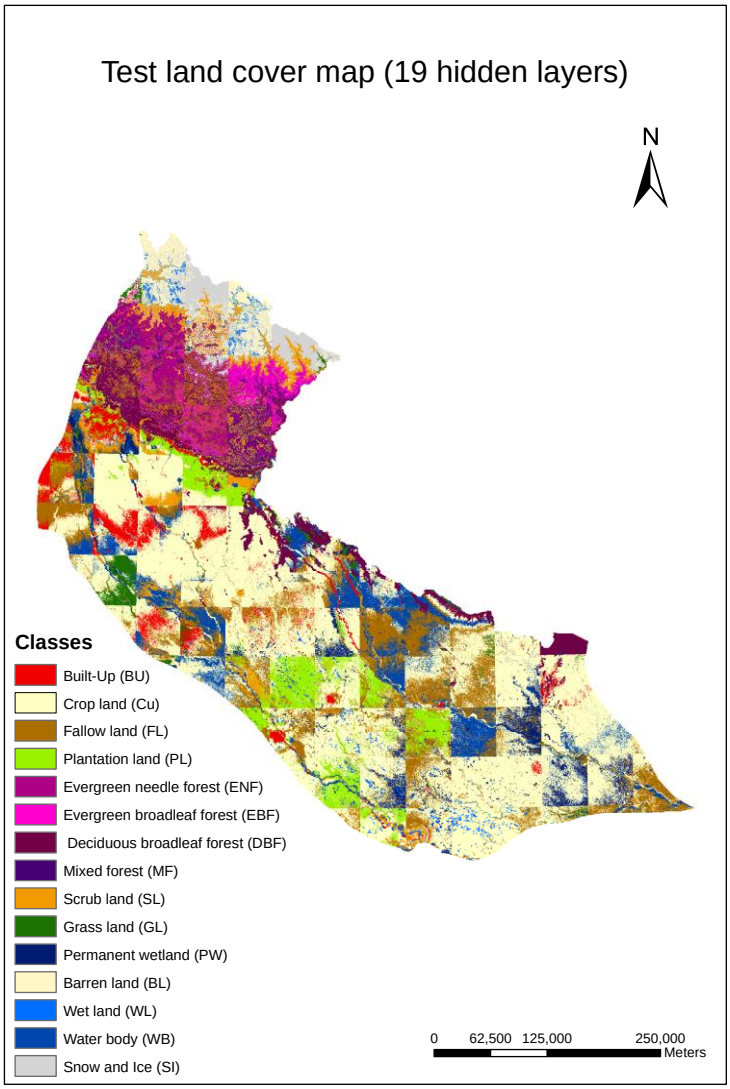
Test maps of Case 2: Table 5-2



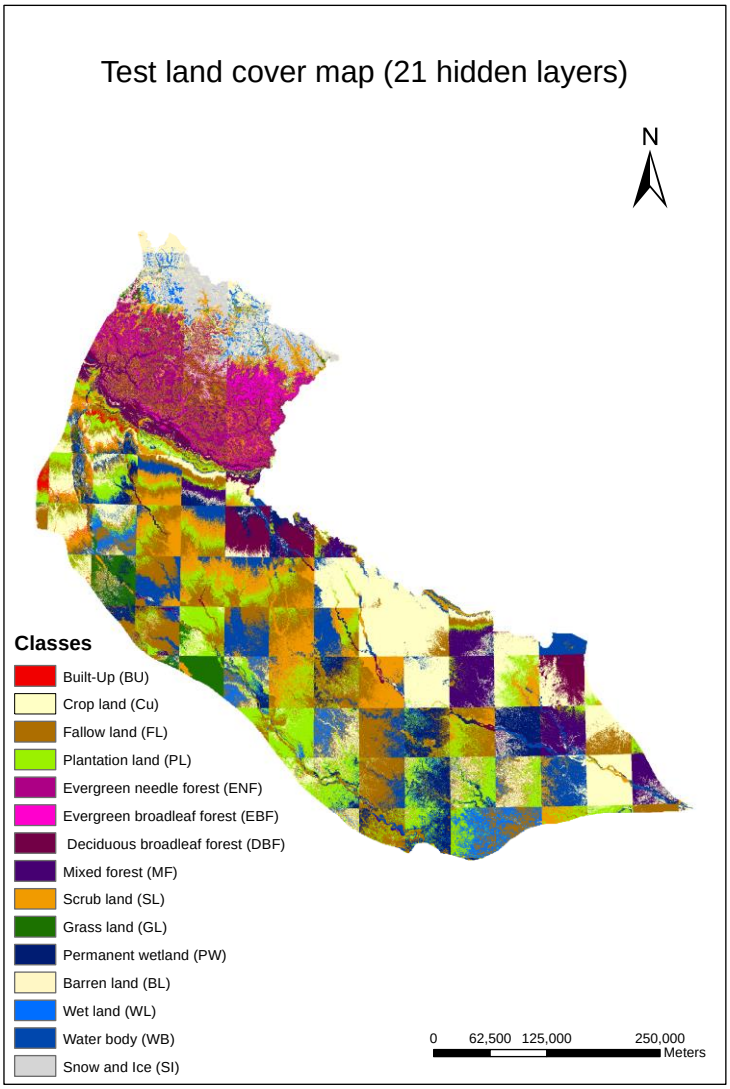
(a)



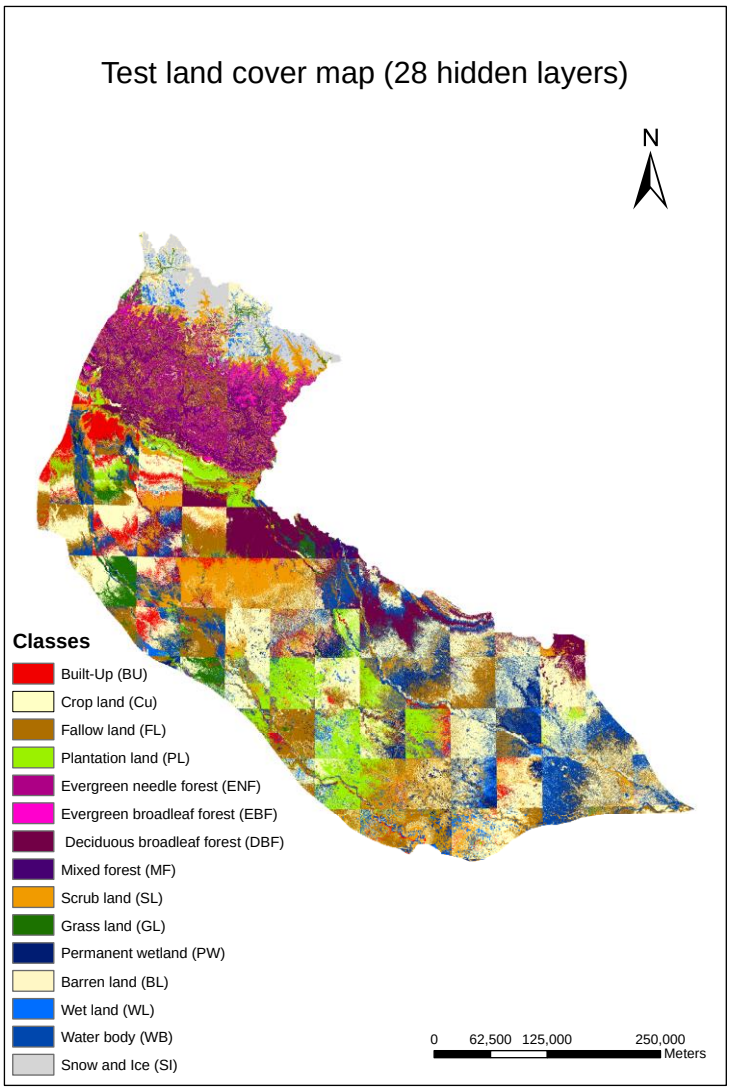
(b)


Test land cover map (19 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

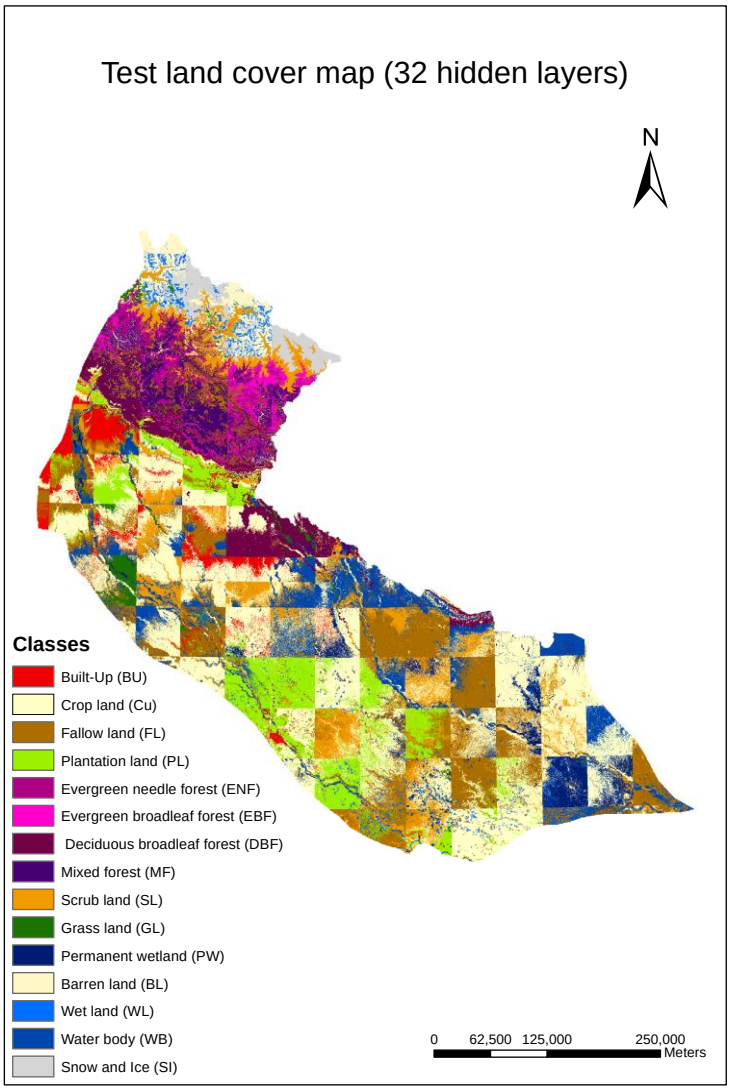
(c)


Test land cover map (21 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

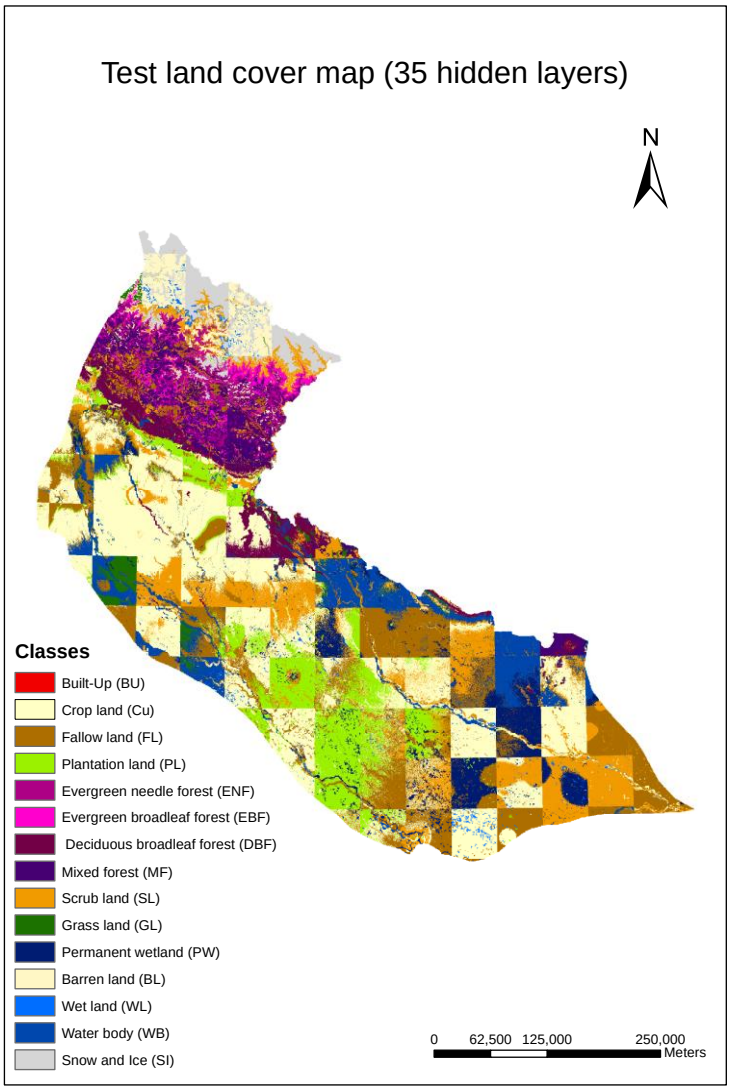
(d)

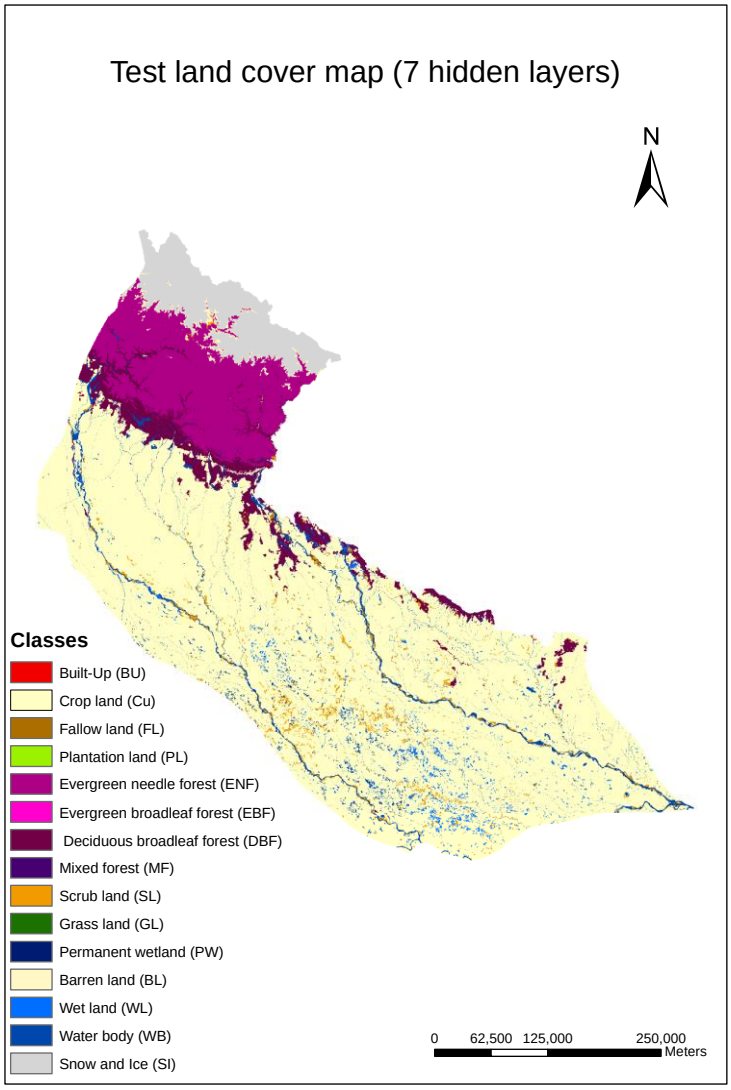

Test land cover map (28 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

(e)


Test land cover map (32 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

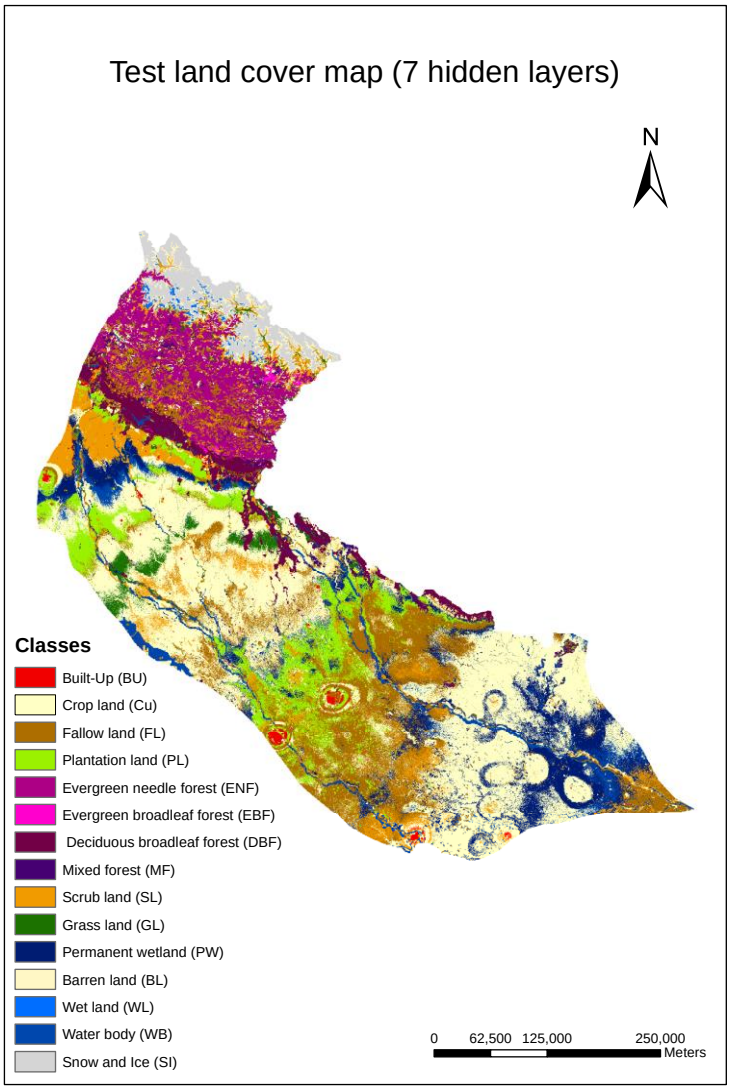
(f)



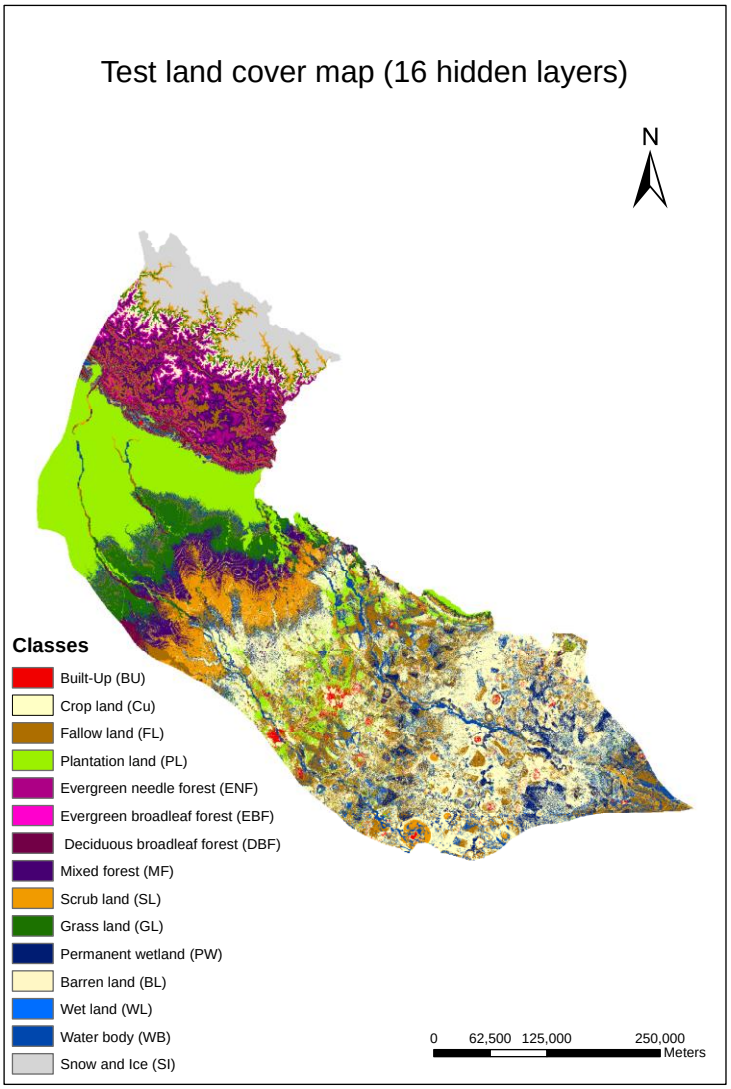


(g) **Test map of Case 2: Table 5-3**

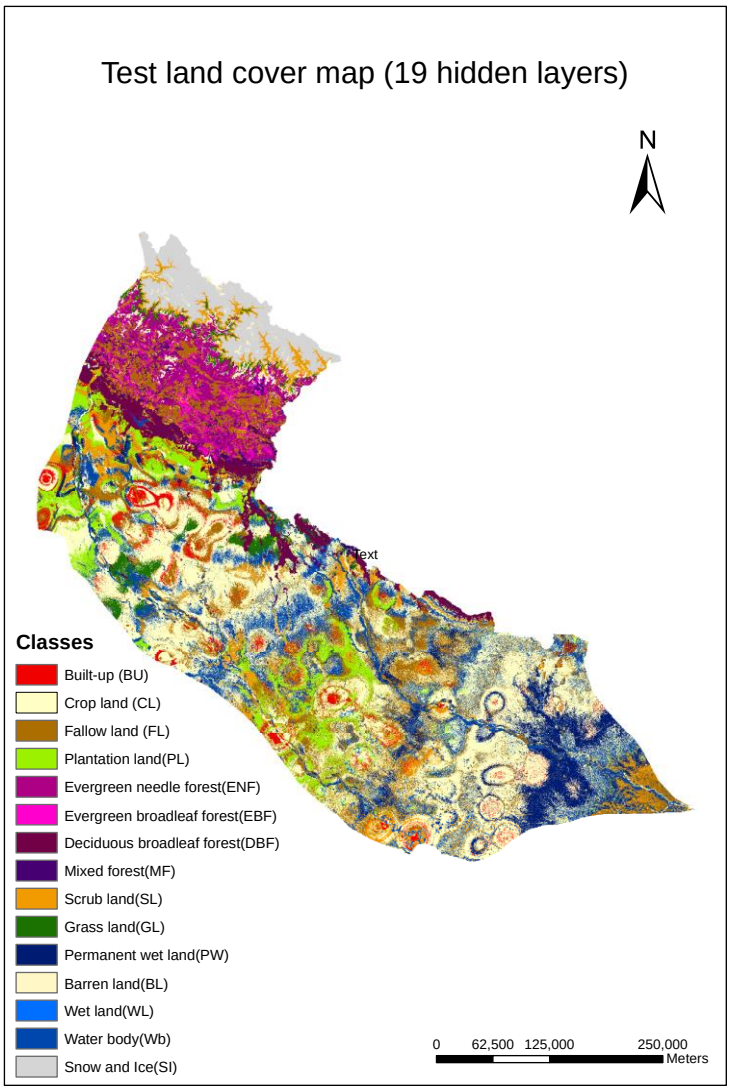
Test maps of Case 3: Table 5-4:



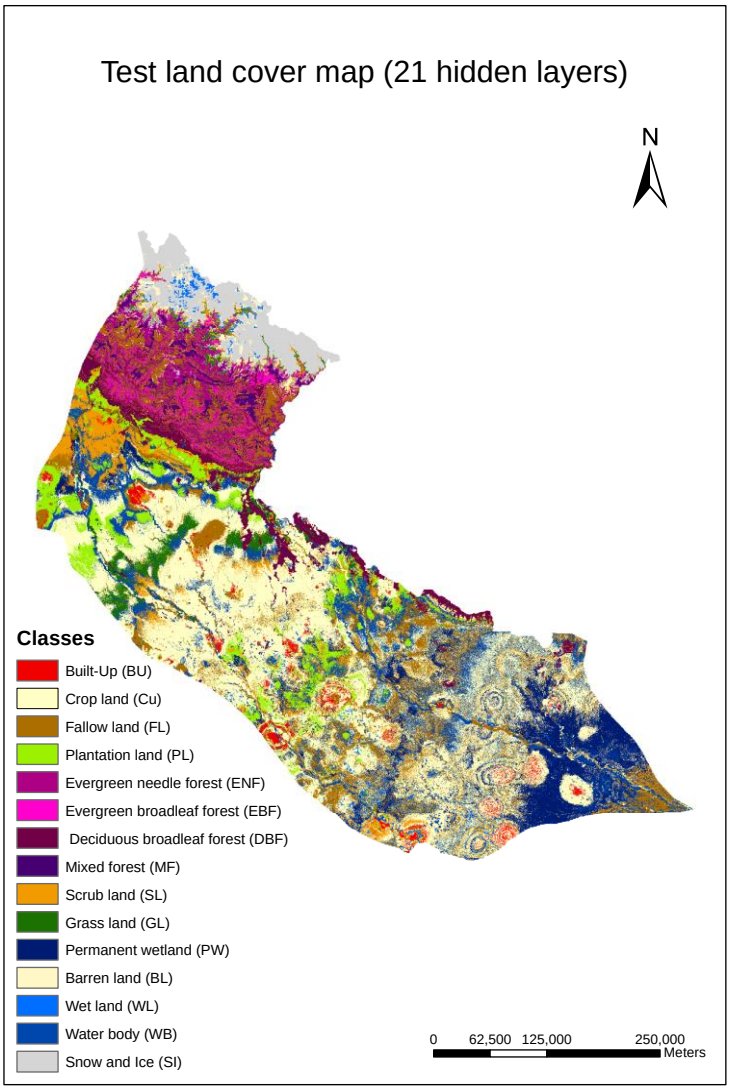
(a)



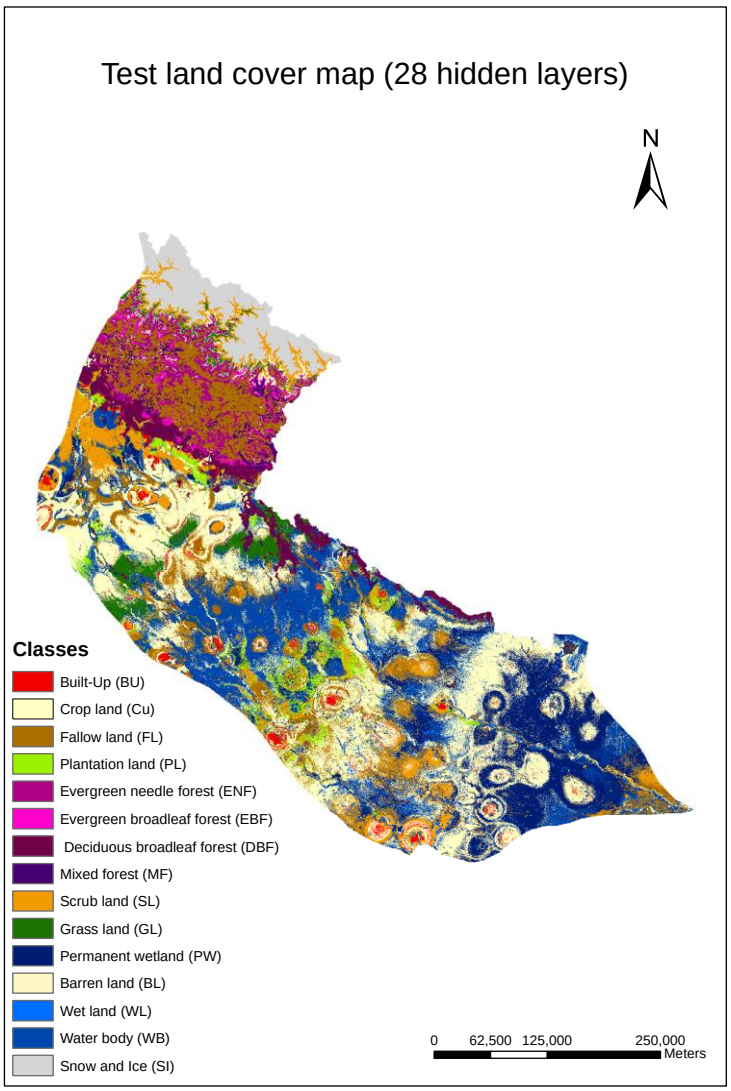
(b)


Test land cover map (19 hidden layers)
N
Classes
Built-up (BU)
Crop land (CL)
Fallow land (FL)
Plantation land(PL)
Evergreen needle forest(ENF)
Evergreen broadleaf forest(EBF)
Deciduous broadleaf forest(DBF)
Mixed forest(MF)
Scrub land(SL)
Grass land(GL)
Permanent wet land(PW)
Barren land(BL)
Wet land(WL)
Water body(Wb)
Snow and Ice(SI)
0 62,500 125,000 250,000
Meters

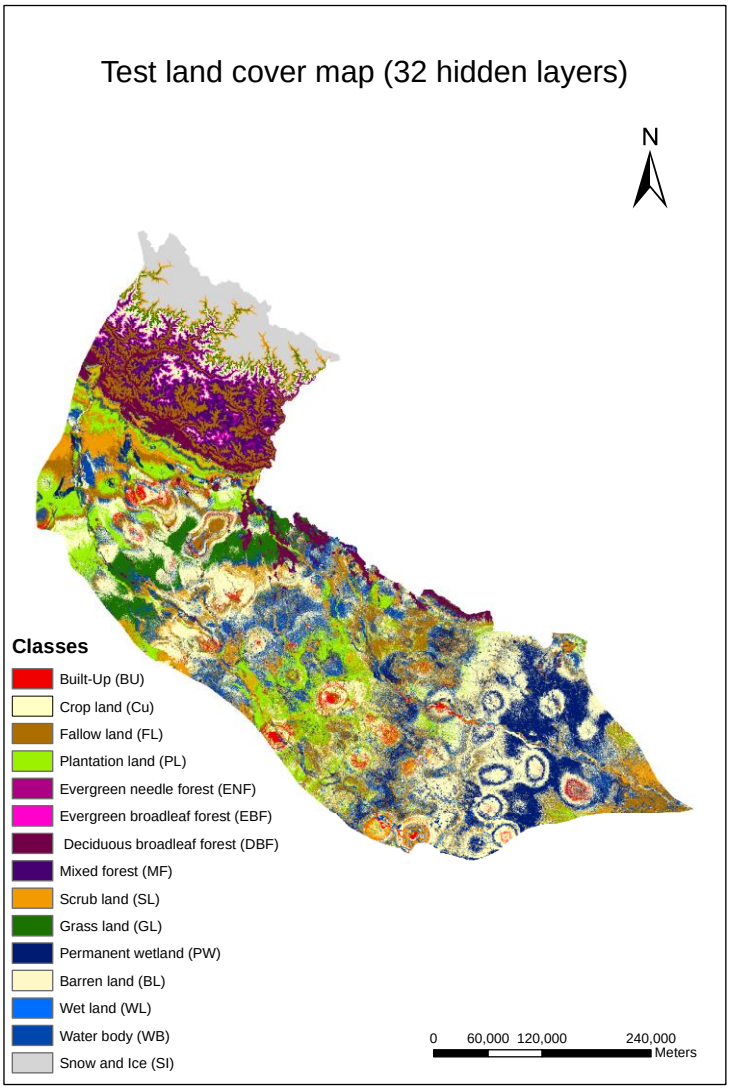
(c)


Test land cover map (21 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

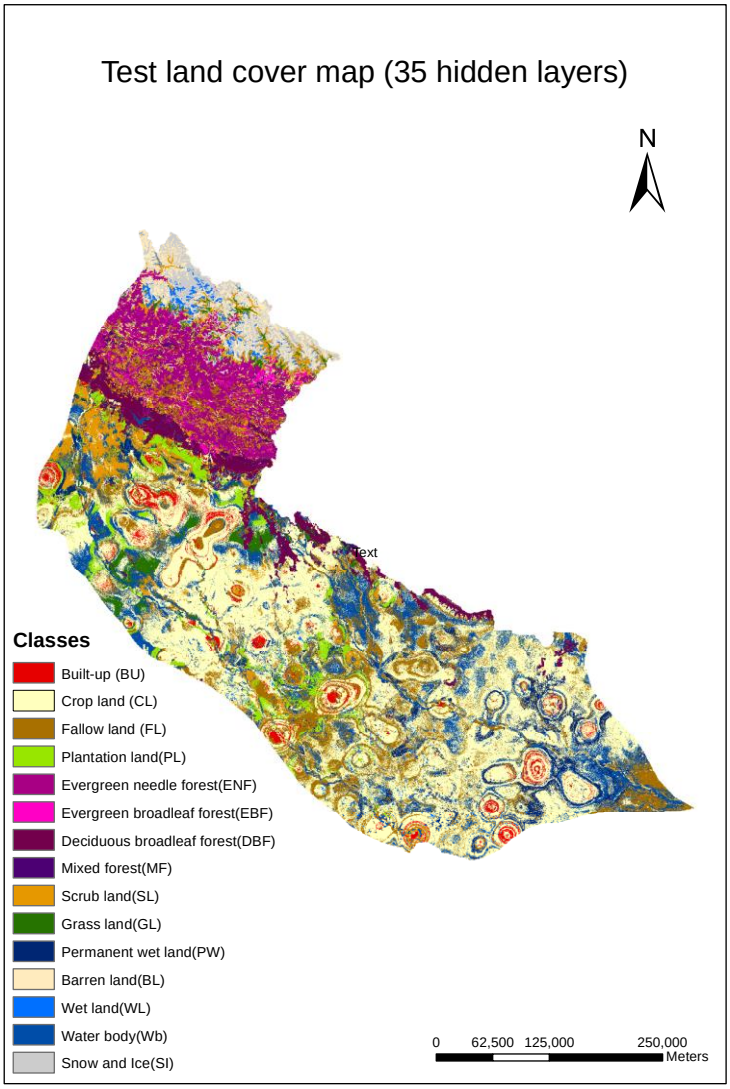
(d)

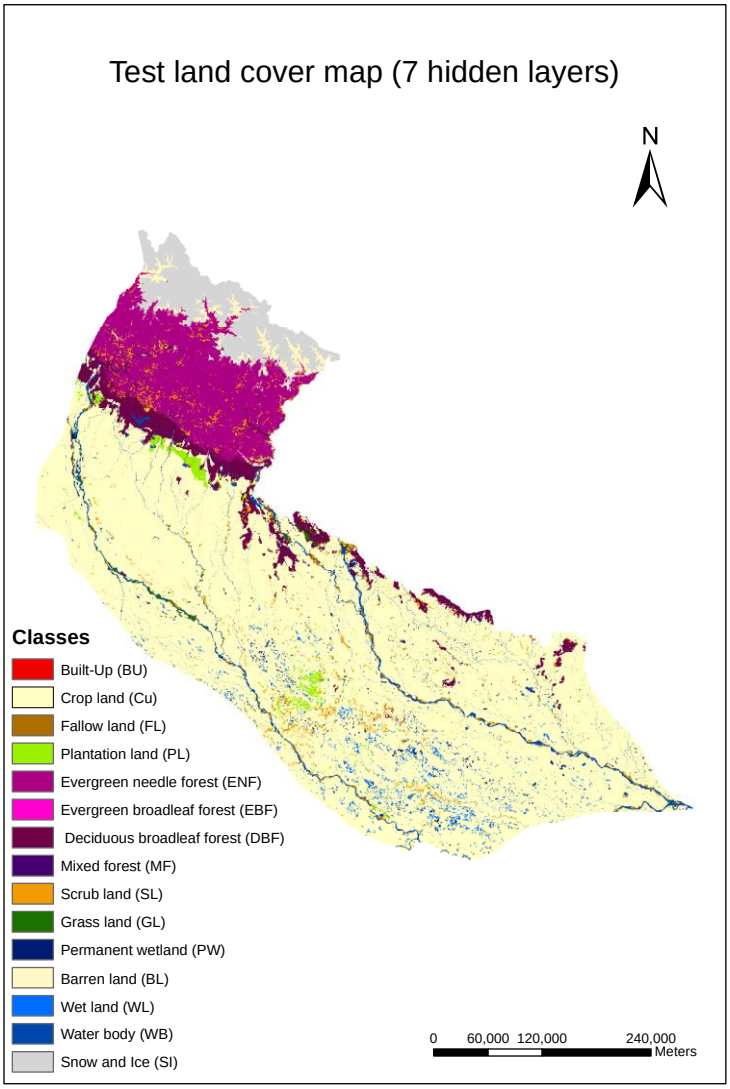

Test land cover map (28 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

(e)


Test land cover map (32 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 60,000 120,000 240,000
Meters

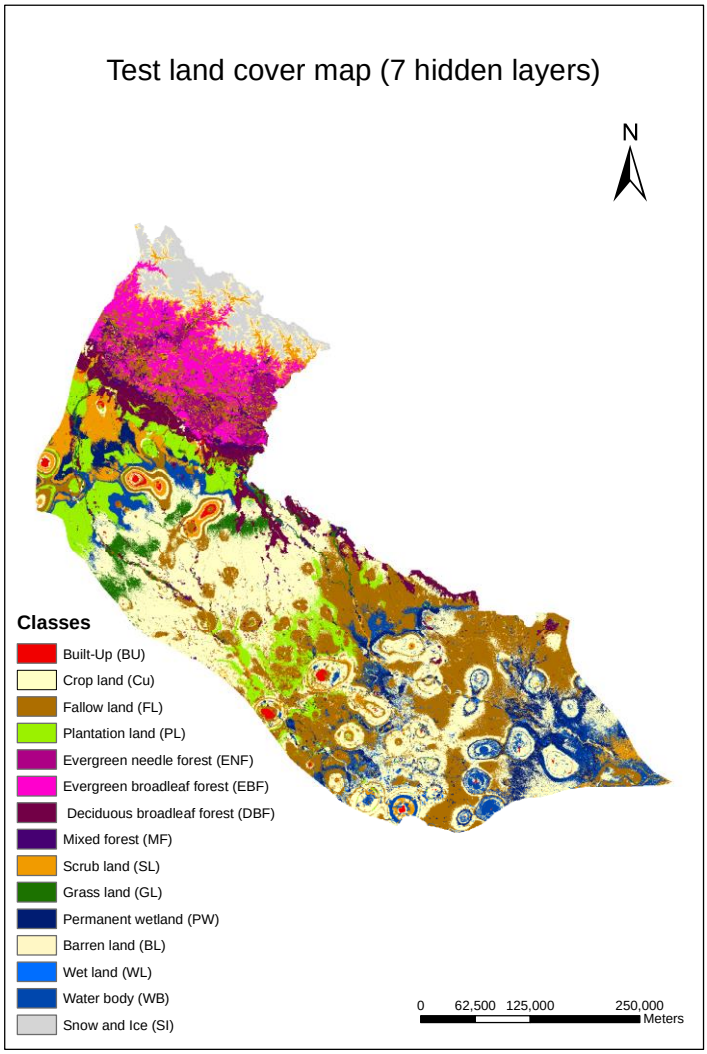
(f)



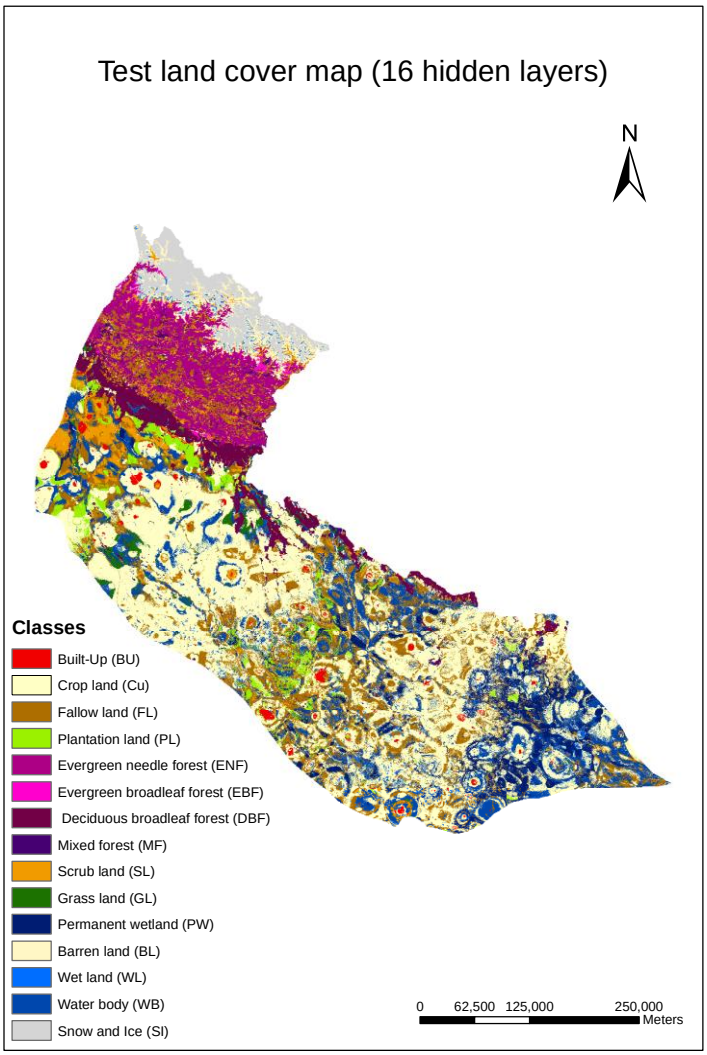


(g) **Test map of Case 3: Table 5-5**

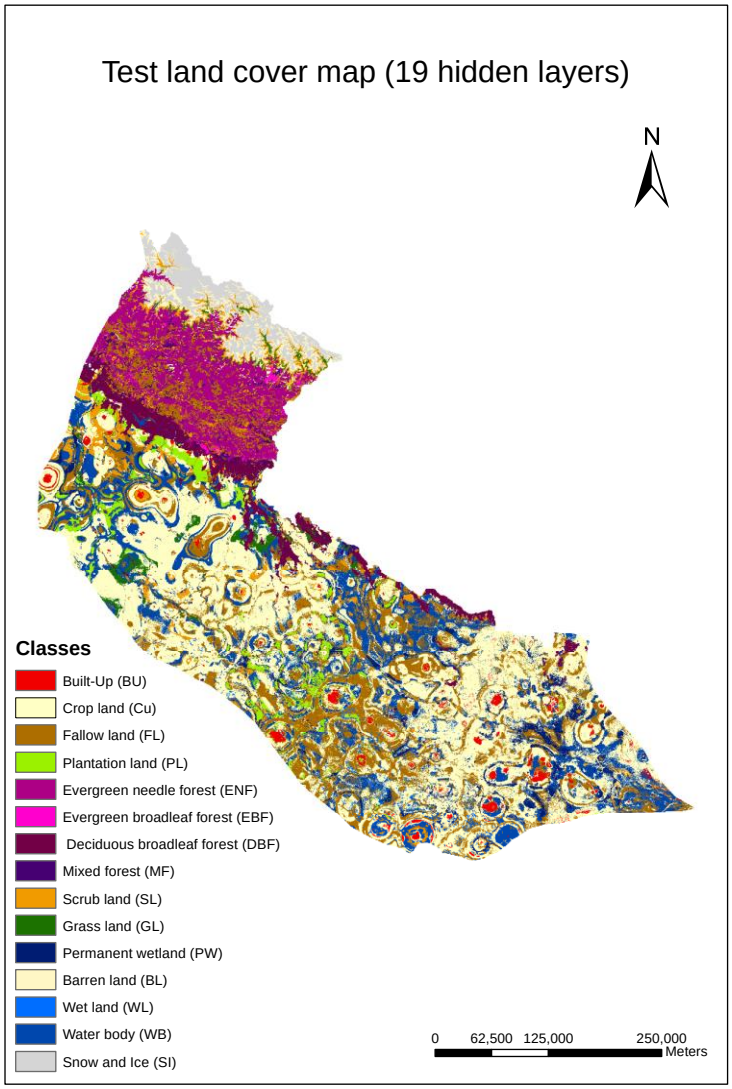
Test maps of Case 4: Table 5-6



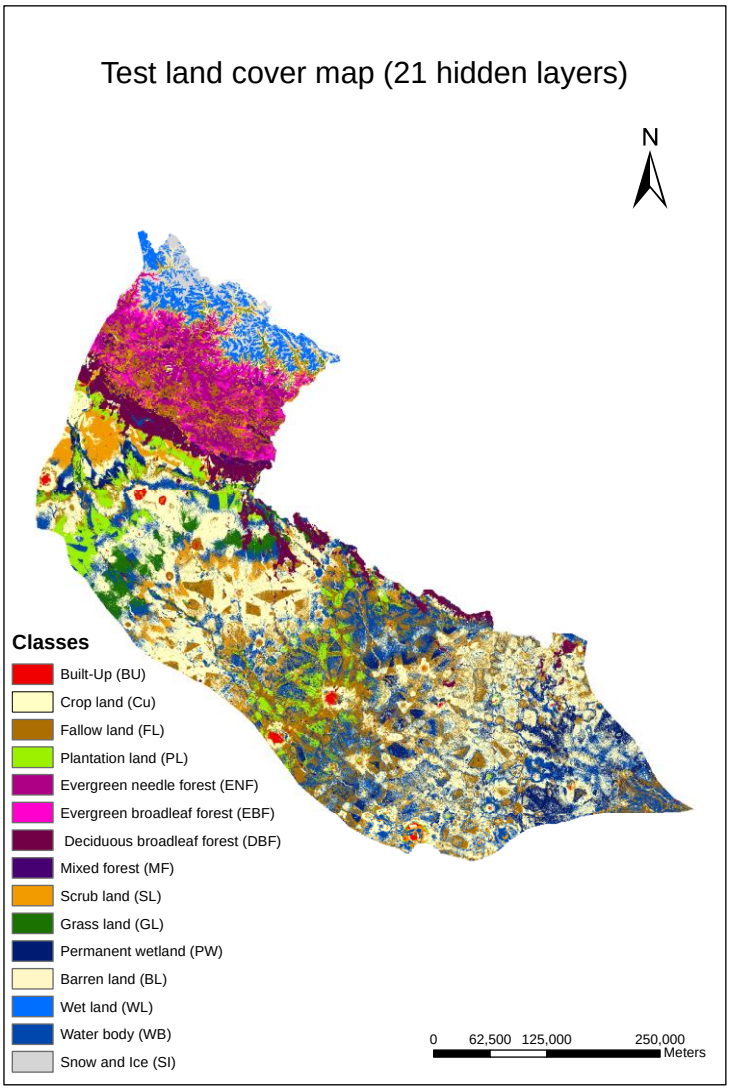
(a)



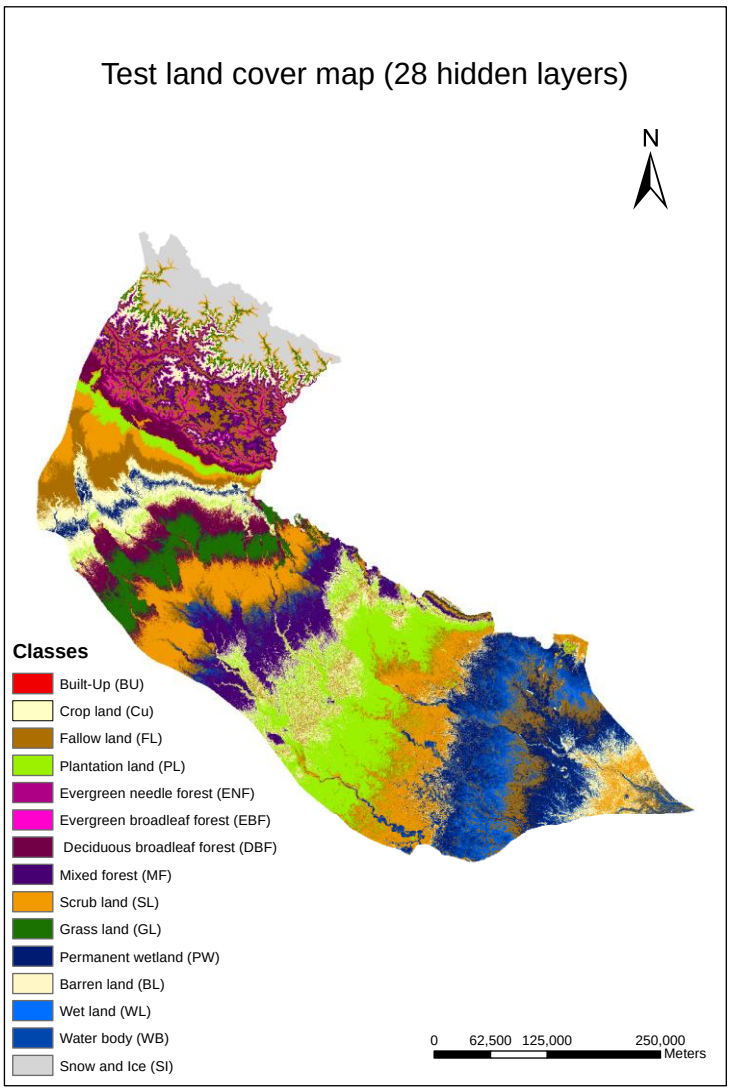
(b)


Test land cover map (19 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

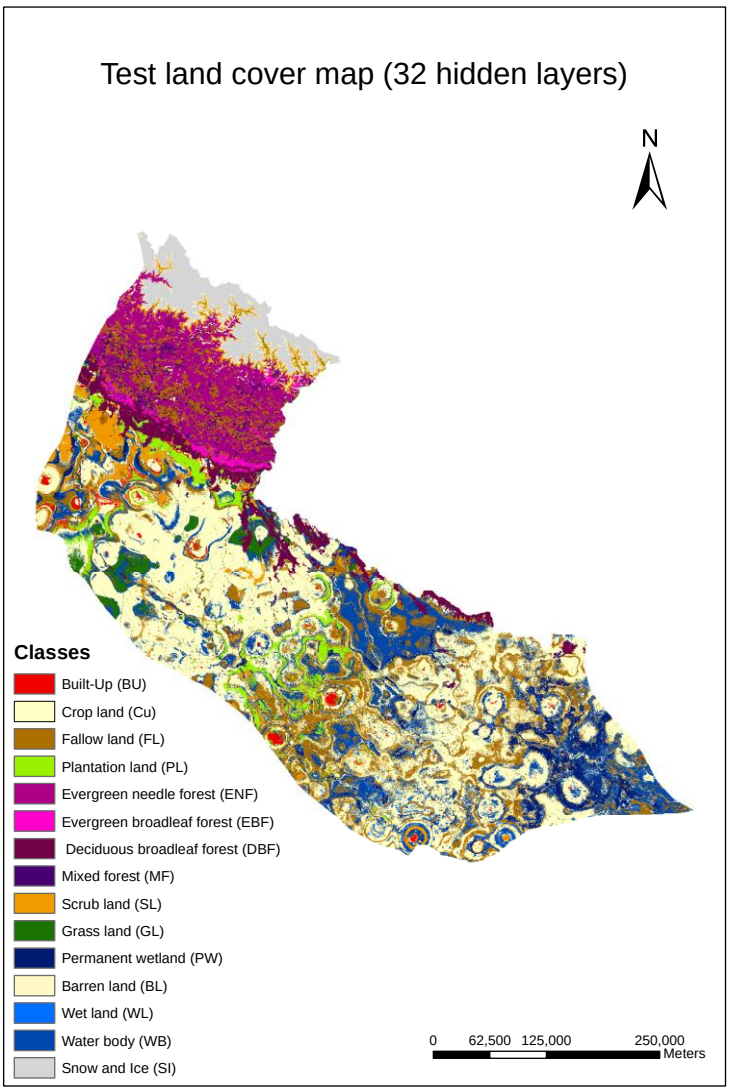
(c)


Test land cover map (21 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

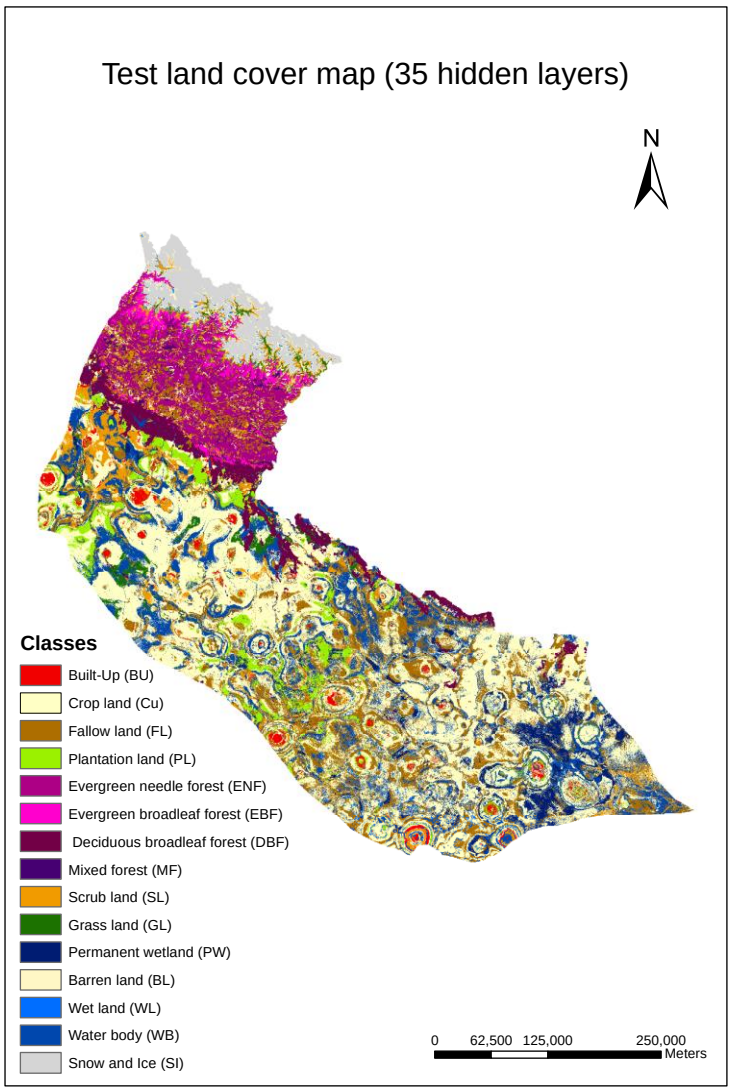
(d)

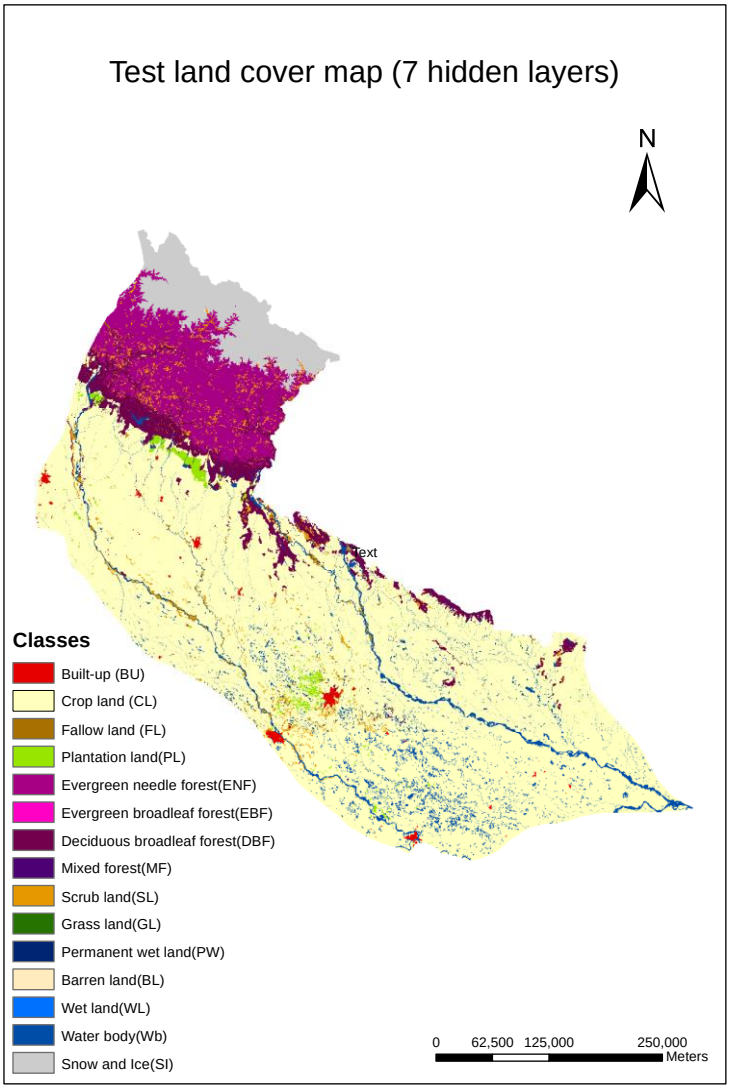

Test land cover map (28 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

(e)


Test land cover map (32 hidden layers)
N
Classes
Built-Up (BU)
Crop land (Cu)
Fallow land (FL)
Plantation land (PL)
Evergreen needle forest (ENF)
Evergreen broadleaf forest (EBF)
Deciduous broadleaf forest (DBF)
Mixed forest (MF)
Scrub land (SL)
Grass land (GL)
Permanent wetland (PW)
Barren land (BL)
Wet land (WL)
Water body (WB)
Snow and Ice (SI)
0 62,500 125,000 250,000
Meters

(f)





(g) **Test map of Case 4: Table 5-7**