Identification and Temporal Forest cover change detection using Machine Learning approach for Sustainable Forest Management; A case study of Mansehra Region, KPK, Pakistan

By

MUHAMMAD TASLEEM KHAN

(2017-NUST-MS-GIS-205342)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Remote Sensing and GIS

Institute of Geographical Information Systems
School of Civil and Environmental Engineering
National University of Sciences & Technology
Islamabad, Pakistan

August, 2021

DEDICATION

"To my loving mother & father who supports me in all moments of life."

&

"To teachers and friends who helps me to complete my thesis…"

ACKNOWLEDGEMENTS

First and foremost, I would like to praise Allah the Almighty, the Most Gracious, and the Most Merciful.

My warm profound thanks and appreciation go to very friendly and enormously cooperative attitude of my respected research Supervisor Dr. Javed Iqbal, who always helped me whenever the need arose. His priceless advice, guidance and energizing debates with him has always been a pleasure and a source of learning for me.

Many thanks to Dr. Muhammad Hassan Ali Baig and Dr. Muhammad Umair who were members of examination committee. They not only extended their full cooperation but also helped me refining and completing this work under their scholarly guidance. I would like to offer my sincere thanks to Dr. Muhammad Shahzad for his guidance and keen interest in this research.

I am also thankful for direct or indirect altruistic support of many whom I haven't mentioned here. I acknowledge all of their sincere efforts.

Thank you everyone.

Muhammad Tasleem Khan

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Explanation
LCLU	Land-cover/land-use
ML	Machine Learning
RF	Random Forest
SVM	Support Vector Machine
CART	Classification And Regression Tree
FCC	False Color Composite
MSE	Mean Square Error
NFT	National Forest Trend
GPS	Global Positioning System
P/A	Perimeter- to- Area
DN	Digital Number
NLCD	National Land-Cover Database
NFT	National Forest Trend
CPEC	Pak- China Economic Corridor
USGS	United States Geographical Survey
OA	Over-all Accuracy
UA	User’s Accuracy
PA	Producer’s Accuracy
ETS	Exponential Tripple Smoothing

ABSTRACT

Forest and wildlife preservation departments in Pakistan face many challenges and are ill-equipped to counter them. Departmental incompetence, ambiguous resource rights, and judicial delays were examples of institutional weaknesses that impede forest monitoring. A novel approach has been devised through remote sensing and machine learning techniques to tackle temporal forest monitoring on a large scale. This study's specific objective was to develop an automatic system to classify and detect forest cover and its trend over 30 years over the Mansehra region, Khyber Pakhtunkhwa (KPK). Machine Learning (Random Forest) algorithm was used to classify and detect forest cover using Free-to-use satellite imageries (Landsat) and datasets. Segmentation was applied to labeled six land-cover classes (2015) and trained (2100 patches) and test (900 patches) classifier (Random Forest (RF)) with an accuracy of 89.47 % and 87.58 %, respectively. Random Forest classifier clearly outperforms the other statistical machine learning methods and is used for the complete analysis of classification (land-cover/ land-use) and forest cover change detection over the Mansehra Region. Training accuracy percentage was calculated for Random Forest classifier, Support vector machine (SVM), Classification and Regression Tree (CART), and Perceptron as 89.47, 66.33, 73.24, and 49.95, respectively. The testing accuracy percentage calculated for RF was 87.58, SVM was 46.81, CART was 78.48, and perceptron was 51.71. The forest cover area detected is 1495.6898 Km^2, 1279.4715 Km^2, 1129.5112 $Km^{2,}$ and 1461.0676 Km^2 for 1900, 2000, 2010, and 2019 respectively. Overall forest cover change (1990-2019) was calculated using 'change detection metrics' with forest gain percentage (9.73%), loss percentage (10.56%), and percentage effective forest change (-0.83%). In few areas (Biari, Batang, and Behri), percentage effective change maybe a surprise and very effective, which means the afforestation in some areas generate overall forest gain percentage in past decades. Areas like Nokot, Kiwai, and Ghanian region's forest effective change percentage were calculated as -15.38, -17.01, and -2.36, respectively. The study's extension might help to get more cross-temporal data so that a more reliable system can be made for continuous monitoring of forest change.

Chapter 1

INTRODUCTION

Forests are considered the lungs of the earth. Forests are a good source of food, shelter, timber, fuelwood (Amir et al., 2018). They enhance water quality, control floods (Sher et al., 2010), have aesthetic value (Wu et al., 2019), source of medicine in the field of medicine (Shiddamallayya et al., 2016), and a key to tourism (Luo et al., 2016), at last, they are important from the sky to the soil. The importance of forests can't be neglected at any cost, and their destruction is alarming for a single community and equally for the whole of humanity (Duguma et al., 2019).

1.1. Background information

Every year, 15 million acres of forests being lost around the world (Curtis et al., 2018). Due to massive agricultural and demographic pressures, the land becomes a precious resource (Pendrill et al., 2019). Some regions of the world's forests were on the verge of becoming wasteland (Allan et al., 2017). Water, shelter, flood, folder, nutrient cycling, cultural, and recreation value were all provided by forests (Dash et al., 2016). Forests also aid in providing wildlife habitat and the mitigation of land deterioration and desertification (Bohn et al., 2017).

Forests were a constantly evolving renewable natural resource. "A modification in the surface components of the plant cover" or "a spectral/spatial movement of a vegetation entity through time" (Milne, 1988) were the terms used to describe the change (Coppin et al., 1996). The rate of change might be drastic and rapid, as in large-scale tree logging, or subtle and slow, as in standing volume growth (Coppin et al., 2004).

1.2. Rationale

On the Earth's surface, there were just a few landscapes still in their natural state. In some way, the Earth's surface is being drastically transformed. The world's forests were being harmed more than ever before as a result of increasing natural and anthropogenic disturbances that had a negative impact on the ecosystem. Every year, 15 million hectares of forest being lost around the world. Due to massive agricultural and demographic pressures, the land is becoming increasingly scarce. Some parts of the world's forests were on the verge of becoming barren ground.

Pakistan has a forest shortfall, with around 4.52-million hectares of land covered in forest, or 5.23 percent of the country's total land cover (Bajwa et al., 2015). The per capita forest is 0.03 hectares as compared with 0.6 hectares of the world level (Saeed et al., 2016). Six main forest types of the country include Conifers 42%, Scrub 34%, Irrigated plantation 6%, Riverine Forest 6%, Mangroves 11% and linear plantation 0.5 % (Champion et al., 1965).

According to the National Forest inventory and survey (2015), under the supervision of the forestry sector master plan (FSMP), the total forest cover area is 3.72 million ha (4.6%) of the total area of Pakistan (Qamer et al., 2016). However, only three regions contribute 69% of total forest cover: Khyber Pakhtunkhwa province 1.57 million ha, Azad Kashmir state 0.32 million ha, and Gilgit Baltistan 0.68 million ha (Dove, 1992). United Nations Food and Agriculture organization observed that every year Pakistan lost about 42,000 ha (1.66%) of trees during a period of twenty years (1990-2010) (MacDicken, 2015). Forest cover is boosted in Khyber Pakhtunkhwa province due to the provincial government initiative of a billion-tree tsunami project in 2013 in the whole province (Sloan et al., 2015). According to forest department records, the total forest cover is in the area raised to a 7.1% growing rate in two-year (1.59 million ha) (Zhang et al., 2020).

In 2009, Pakistan's wood consumption was predicted to be 4.42-million-meter cubes, while total wood output was 3.93-million-meter cubes, showing a 0.437-million-meter cube disparity between consumption and production (Zeb et al., 2019). Thus, the area under forest cover is insufficient to meet the country's demand for wood, fuelwood, and wood-based products (Khan et al., 2020).

For development initiatives, policymakers, and environmental organizations attempting to enhance forest management, forest change poses significant obstacles. These difficulties can arise naturally or as a result of local communities living near a forested area. As a result of this relationship, the natural forest cover faces a major threat from the growth of agricultural requirements.

Urbanization, farming, overgrazing, global warming, and tourism development were the primary causes of deforestation. Northern Baluchistan's juniper woods had been intensively plundered for lumber and fuelwood.

Ecological shifts in the floodplain of the Indus River had a significant impact on riverine forests. Large swaths of land had been cleared for farming. Logging for lumber and firewood, as well as clearings for agriculture and human settlements, were all putting a strain on the Himalayan temperate forests. Pakistan's deforestation rate is rising at a rate of 0.2 to 0.5 percent per year.

More ever, local needs for fuel, wood, lumbering, and illegal trade add to the forest cutting. Such rapidly decreasing forest cover on the hand and slow generation of natural vegetation makes it unfavorable for forest conservation and extension in this area. To cope with these challenges, forest management must stop deforestation and forest rehabilitation in mountain areas as a prerequisite, mapping has remained the best tool for long in forest management. Nowadays, forest mapping has got special consideration worldwide due to its importance for forest conservation and better management, which would ultimately improve the environment and economy of a country.

1.3. Forest monitoring using GIS and remote sensing techniques

For development initiatives, legislators, and environmental organizations aiming to enhance forest management, forest change poses considerable problems. (Lazdinis et al., 2019). These challenges can both be natural and created by local communities living nearby the forest habitat. Such an interaction caused a severe risk to the natural forest cover in spreading agricultural needs as a sequel. More ever, local needs for fuelwood, lumbering, and illegal trade add to the forest cutting. Such rapidly decreasing forest cover, on the other hand, and the slow generation of natural vegetation make it unfavorable for forest conservation and extension in this area. To cope with these challenges, forest management must stop deforestation, and forest rehabilitation in mountain areas as a prerequisite mapping has remained the best tool for long in forest management (Zhu et al., 2019).

Nowadays, forest mapping has got special consideration worldwide due to its importance for forest conservation and better management, which would ultimately improve the environment and economy of a country (Schulze et al., 2019). For this purpose, the state of the modern art technologies such as Remote Sensing and Geo-information system is used for LCLU and forest cover change (Herold et al., 2019). A variety of methods for analyzing multi-temporal satellite pictures can be used to detect forest changes. (Katarki et al., 2019). The development of vegetation mapping has drastically boosted LCLU change, allowing for an accurate assessment of the world's forest extension and health. Grassland resources had become a top priority as a result of this improvement. (Park et al., 2019). Understanding the impact of man's activities on his natural resource base over time requires viewing the Earth from space.

Observations of the earth from orbit provided objective information of human use of the landscape in rapid and frequently undocumented forest change. Data from Earth sensing satellites has been increasingly important in charting the Earth's features and forest changes and managing natural resources in recent years (Niraula et al., 2013).

1.4. Machine learning algorithms for land –cover classification

State-of-the-art machine learning algorithms were very popular these days. They can aid in forecasting of the change in forestation and analysis of forest cover trends is important because of environmental crisis. However, the Pakistani forestation department did not have state of the art system to detect the changes in forestation and very less data regarding the forestation or deforestation rate in Pakistan is available online. The perceptron is a supervised learning technique for binary classifiers in machine learning. A binary classifier is a function that can determine whether or not a vector of numbers representing an input belongs to a given class.

The Random forests are an example of supervised, regression, and other tasks that works by training a large number of decision trees and then outputting the class that is the mode of the classes (classification) or the mean prediction (regression) of the individual trees. Decision trees had a tendency to over fit their training set, which is corrected by random decision forests.

A Support Vector Machine (SVM) is a method for classifying objects based with a separating hyperplane as its formal definition. In other words, the algorithm produces an ideal hyperplane that categorizes fresh samples given labeled training data (supervised learning).

1.5. Forest cover change detection

As the largest terrestrial ecosystem on the planet, forests were affected by climate change and mankind's activities. In turn, forests influence local, regional and global climates through physical, chemical and biological processes. The study of land-cover and land-use change had spurred a significant interest in the research of forest cover change and its climate influence. Because it provides a plethora of geographical and temporal information, remote sensing is a great tool for assessing changes in forest cover.

The study uses time-series data of Landsat satellite imagery (1990-2019) for approximately 30 years for forest cover change detection and trend analysis. The methodology adopted in this study could be used in other parts of the world, especially in those regions where forest monitoring on regular intervals is not feasible due to dysfunctional forest institutions.

1.6. Study objective

The main objectives of the study were Spatio-temporal analysis of current and future forest cover projection of Manshera Region using time series data (30 years).

The specific objectives of the study were as follows:

- Comparison of different Machine Learning techniques (Random Forest, Support Vector Machine, CART classifier and Perceptron) for automatic classification/segmentation of forest cover of the study area.

1.7. Literature review

Liu et al. (1993) digitized land-use maps of 1934, 1988, and 1941 and generated a GIS database. These maps were then evaluated to identify deforestation rates and their relationships with parameters including forest distance from roads and forest fragmentation, as assessed by the perimeter-to-area ratio (P/A ratio) of forest patches for this country. The data show that between 1934 and 1988, the Philippines lost a total of 9.8 million hectares of forest. The existence of major roadways had a significant impact on deforestation. The higher the rate of deforestation, the closer the forest was to the roadways. By 1988, over 78 percent of the 2.1 million acres of forest within 1.5 km of roadways in 1934 had been cleared. Between 15.0 and 16.5 kilometers from highways, however, just 39.5 percent of the forest was lost. The distance between forest patches and highways and the density of roads per unit area did not indicate deforestation. The P/A ratio of forest patches was also helpful in determining their clearing pace. The higher a forest patch's P/A ratio, the more likely it was to be cleared. By 1988, all forests with P/A ratios greater than 65 m per ha had vanished. Small areas and nearby agricultural fields were both characteristics of forests with high P/A ratios in 1934.

For mapping in the forest region, a time and cost-effective way of exploiting low-resolution satellite data were used in series to detect changes in forestation region in Karnataka, India. It employed LANDSAT MSS data Landsat false-color composites to distinguish between 6 different forest types: evergreen, semi-evergreen, moist deciduous, dry deciduous, degraded, and scrub forest. Forest and non-forest regions were identified in the two periods of imaging. Color, tone, texture, shape, size, shadow, position, and association were all interpreted, and categories were determined using the "convergence of evidence" approach. The millimeter grid was used to determine the area. A stratified random sampling methodology was used to measure map accuracy, yielding an accuracy of 85.5 percent (Kushwaha, 1990).

Khan et al. (2017) studied that at a sub-annual level, forest monitoring is carried out utilizing the automatic solution. Detect changes in forest cover over a 29-year period (1987–2015), despite the fact that the underlying data is woefully poor and contaminated with artefacts. To begin, the lost surface reflectance information is recovered using a spatiotemporal in painting process. Following a sparse encoding-based reconstruction, the spatial filling procedure employs the available data from surrounding temporal instances. Forest modifications were determined and their onset and offset times were predicted by labelling the candidate set of suggestions using highly discriminative representations. The unconstrained chance discovery in a large geographic area handles multi-class change detection and estimation of the start- and end-times of the detected change event by selecting class independent change event feature vectors and anticipating the likelihood of certain change event types, as well as their start- and end-times. This method obtained a state-of-the-art average patch classification rate of 91.6 percent (a 16 percent improvement) and a mean onset/offset prediction error of 4.9 months (a 16 percent improvement) (an error reduction of five months). With the ability to assess the identified changes in the unlabeled image region qualitatively.

Afify (2011) worked on post-classification, image differencing, image rationing, and principal component analysis for change detection in areas of Egypt. To achieve a real change, the grey levels of images of the same area must be almost same, and the technique of histogram equalization must be applied. Then, using maximum likelihood (ML), each pixel was labelled into one of three classes, and the bi-temporal images were pixel by pixel compared to generate six bitmaps reflecting the modified and unchanged areas. The accuracy was 66.7 percent when compared to reference maps provided by the Egyptian general survey authority. Change was recognized in image differencing by comparing digital numbers in bi-temporal images and deciding the change based on a pre-determined SD value. The accuracy of this approach was 53.56 percent. The ratio of DN values is compared with the threshold determined based on the set SD value in picture rationing.

Xian et al. (2009) applied radiometric change detection of probable land cover changes from paired satellite images to update the National Land Cover Database (NLCD). The linear regression algorithm compares each pixel of the object image, which is usually a late-date image that will be altered, to the reference image, which is an early-date image that will be left alone. A change vector image depicting spectral feature difference and representing changes in land-cover is calculated using the normalized and reference pictures. An urban mask is then created using the revised land cover. The 2006 impervious surface is then updated using a regression tree model trained from unaltered pixels and a baseline of the NLCD 2001 impervious surface product.

Lehmann et al. (2013) used operational procedures to produce, National Forest Trend (NFT) data, which is a time-series summary with a visual indication of within-forest vegetation changes over time at a resolution of 25 meters. A consistent proxy for tracking density changes is a time series of index values for a location produced from calibrated images. The varied colors represent the approximate timing, direction (decline or increase), magnitude, and spatial extent of the changes in vegetation cover.

Over the year many studies (Qamer et al., 2012), (Hubacek et al., 2013), (Joshi et al., 2008) and (Qamer et al., 2016) conducted throughout the world to study the forest cover change using remote sensing data utilizing traditional analysis techniques. However, this study presents a novel technique of forest cover change, which automates the analysis technique through machine learning using the random forest algorithm. Furthermore, the study uses time-series data of Landsat satellite imagery (1990-2019) for approximately 30 years for forest cover change detection and trend analysis.

Chapter 2

MATERIALS AND METHODS

2.1 Study area

The area of study, Mansehra region (34º - 14' to 35º - 11' N and 72º - 49' to 74º - 08'E), is situated in the North-East Himalayan ranges of Khyber Pakhtunkhwa (KPK) province (Figure1) in Pakistan. It consists of five tehsils and the total area is 4,153.5293 km^2. Forest and agriculture were the primary natural resources of this region (Awan et al., 2011). The region lies in the Himalayan region with an elevation range from 400-5217 meters. The slope, aspect, rainfall, temperature, and humidity are the region's most critical environmental factors (Awan et al., 2011; Naz et al., 2011). The increasing population (2, 39,745 in 2018) and infrastructure development change the Mansehra region's importance and affect the forest and other natural resources (Khalid, 2020).

Karakorum highway, the most important route for Pak-China economic corridor (CPEC) passes through this district. According to the provincial forest department KPK, the forest cover area's total was 1296.4656 km^2 (129646.56 ha) in 2015. The dominant forest species are *Pinus roxburghii* (chir pine), *Pinus wallichiana* (blue pine), *Cedrus deodara* (deodar), and *Abies pindrow* (Fir) (Hussain et al., 2016). Naran and Kaghan consist of even and uneven forests. They were considered the most beautiful tourist spots in Pakistan (Khan et al., 2018). Rainfall is controlled by two factors, namely moon-soon winds and elevation. The surface configuration is generally very rugged, having uneven slopes.

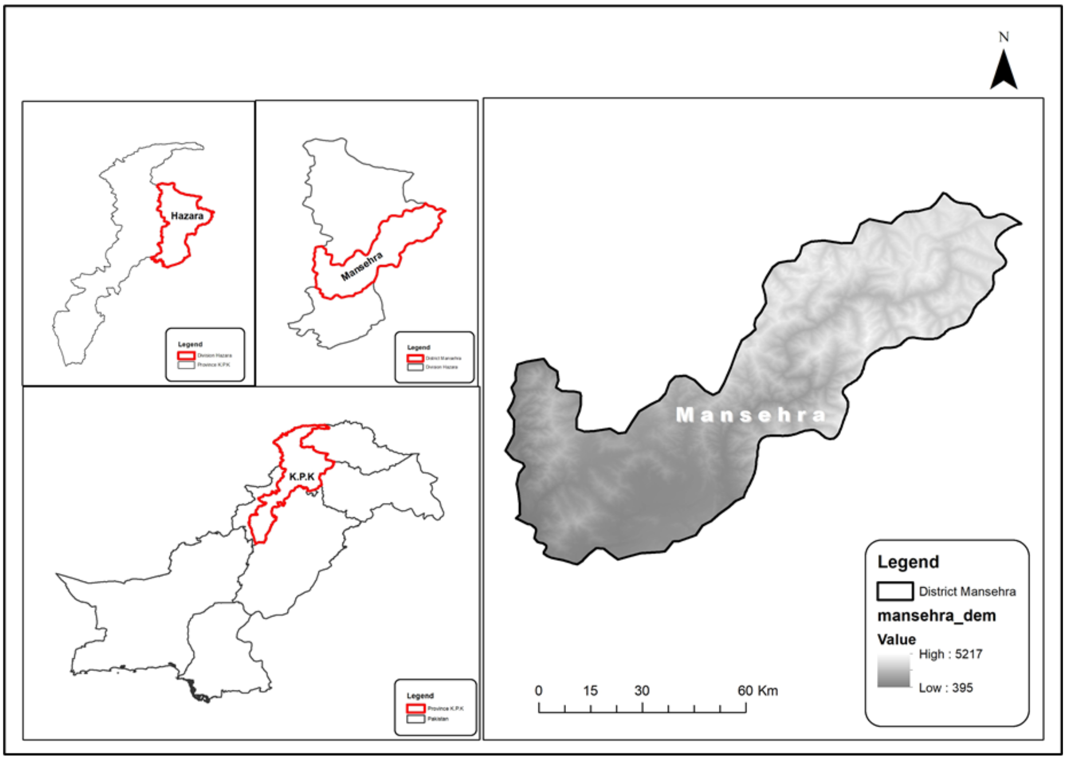


Figure 2.1. Showing the map of Pakistan, Khyber-Pakhtunkhwa, Hazara division and the study area Mansehra Region.

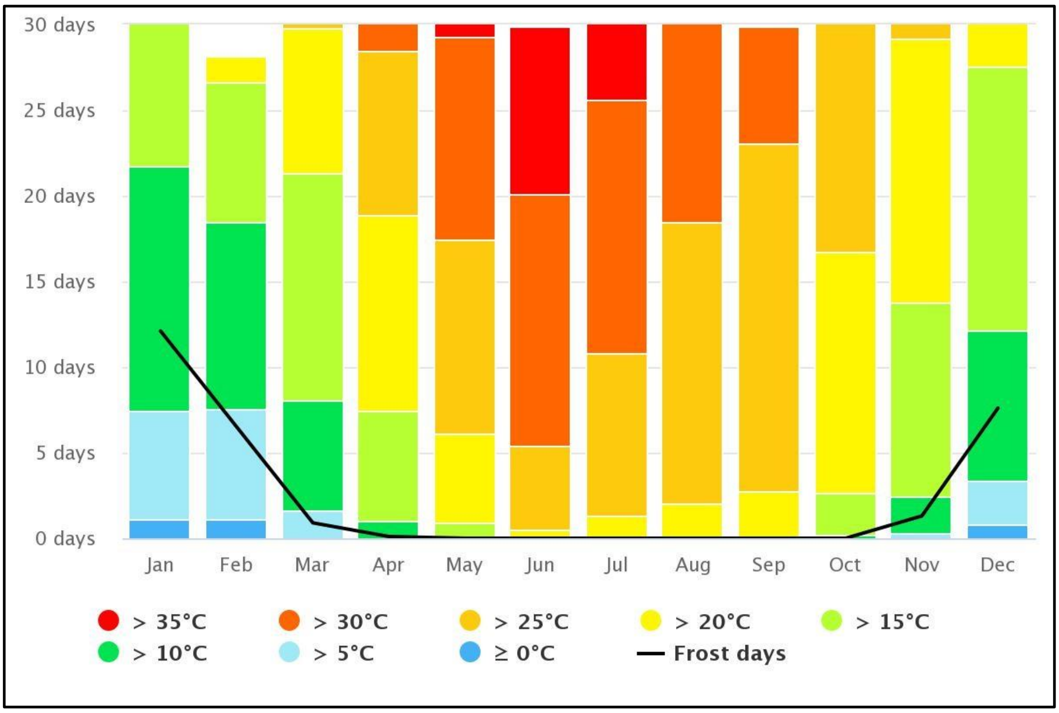


Figure 2.2. Showing daily (average) temperature of Mansehra Region for year 2019. Source: (timeanddate.com).

2.1.1. Population

The population was 1,556,460 people according to the 2017 census. In the 1998 census, the total population of the Mansehra, Balakot, and Oghi tehsils was 978,200. It was characterized by high mountains, which ranged in elevation from 200 meters above sea level in the south to about 4500 meters in the north.

2.1.2. Climate and topography

Average temperature is about 19 C^o. Hottest month of year is June and July having an average temperature of 27 C^o while the coldest month is January having average temperature 9 C^o. The average annual precipitation is 1195mm. The driest month is November with 32mm of precipitation and the wettest month is August having 270 mm of average precipitation.

Rainfall is controlled by two factors, namely moon-soon winds and elevation. Its elevation is about 5200m having various topographic features such as valley, ridge, and hill shade. The surface configuration is generally very rugged, having uneven slopes.

Heavy snowfall occurs in. Snowfall starts from December and stay till April in upper kaghan. Naran and Sirin valley, but for the last few years, the timing and duration of rainfall and snowfall, altogether had been changed, causing the climatic shift. Average snow depth varies from one to three feet depending upon weather conditions and height from sea level. Severe frost is received in plains, particularly along the watercourses.

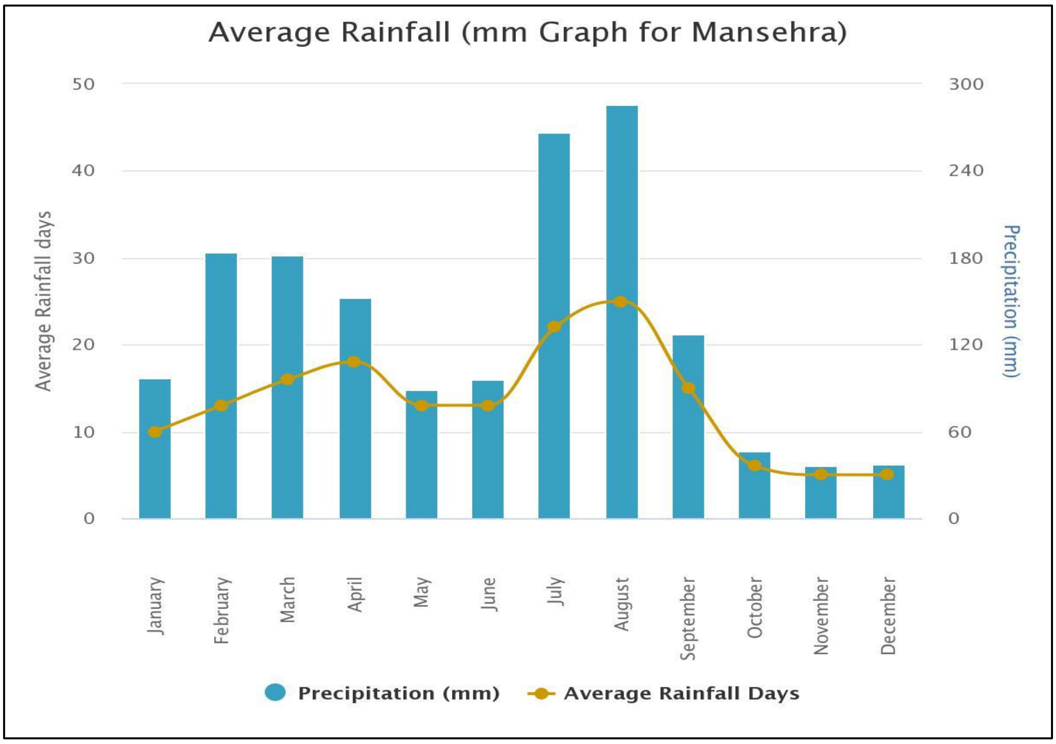


Figure 2.3. Showing average rainfall days and precipitation (mm) of the Mansehra Region for the year 2019. Source: (timeanddate.com).

2.1.3. Flora and fauna

The main plant species found in the Mansehra region are Chir pine (*Pinus roxburgi)*, Deodar (*Cedrus deodara)*, Spruce, Fir, and other evergreen species. On steeps of upper Kaghan bushes of Pushtun juniper *(Juniperus macropoda)*, Dwarf juniper *(Juniperus communis)* as well as Birch (*Betula utilis)* trees are found. The Fauna of area consists of snow Leopard (*Panthera uncia*), Brown Monkey (*Ateles hybridus*), Urial (*Ovis vignei*), Wild Goat (*Capra aegagrus*), Chir Pheasants (*Catreus wallichii*), Bar-tailed pheasant (*Syrmaticus reevesii*) and other migratory birds.

2.2 Materials

Continuous monitoring is necessary to check the human activities which directly affect biodiversity (Zhu et al., 2020). Most researchers and scientists used traditional forest data collection methods, which consist of forest inventory, measure each individual tree, plot method inventory, and prism method survey (Almeida et al., 2019). These methods need most of the capital and laborious work; therefore, new techniques were used in satellite images and drone images in the modern world (Almeida et al., 2019). To avoid problems associated with traditional data collection methods, remotely sensed datasets were used in this research to identify the forested area and temporal changes of the past 30 years using machine learning classifiers. Data extraction of forest region was the most important step for training and testing an intelligent system to detect and predict the change in forest cover of the area of interest. Two sets of data for both leaf-on and leaf-off seasons, Landsat images were obtained by downloading from the United States Geological Survey (USGS)

Table 2.1. Datasets used in this study.

Data	Description	source
Landsat imageries	Both leaf-off and leaf-on seasons for year 1990, 2000, 2010 and 2019	United States Geological Survey (USGS) https://earthexplorer.usgs.gov)
Geographical Topo-sheet of Pakistan (2003)	Mansehra Region boundary	Survey of Pakistan, Map sale point, Rawalpindi, Pakistan
DEM data	30-meter resolution (Slope, Aspect)	United States Geological Survey (USGS) https://earthexplorer.usgs.gov)
Referenced forest count data	Forest canopy cover area, number of blocks/plots	REDD++ (Pakistan) and Forest departments Mansehra and Upper kaghan
Referenced forest cover labelled images	Land-use/landcover Classified image for year 2015	Forest department KPK and Office of Billion tree tsunami project Peshawar
Rainfall	Average annual rainfall for year 1990, 2000, 2010 and 2019	The Tropical Rainfall Measuring Mission (NASA) http://trmm.gsfc.nasa.gov
Temperature	Average Annual Temperature for year 1990, 2000, 2010 and 2019	World Climate data https://worldclim.org

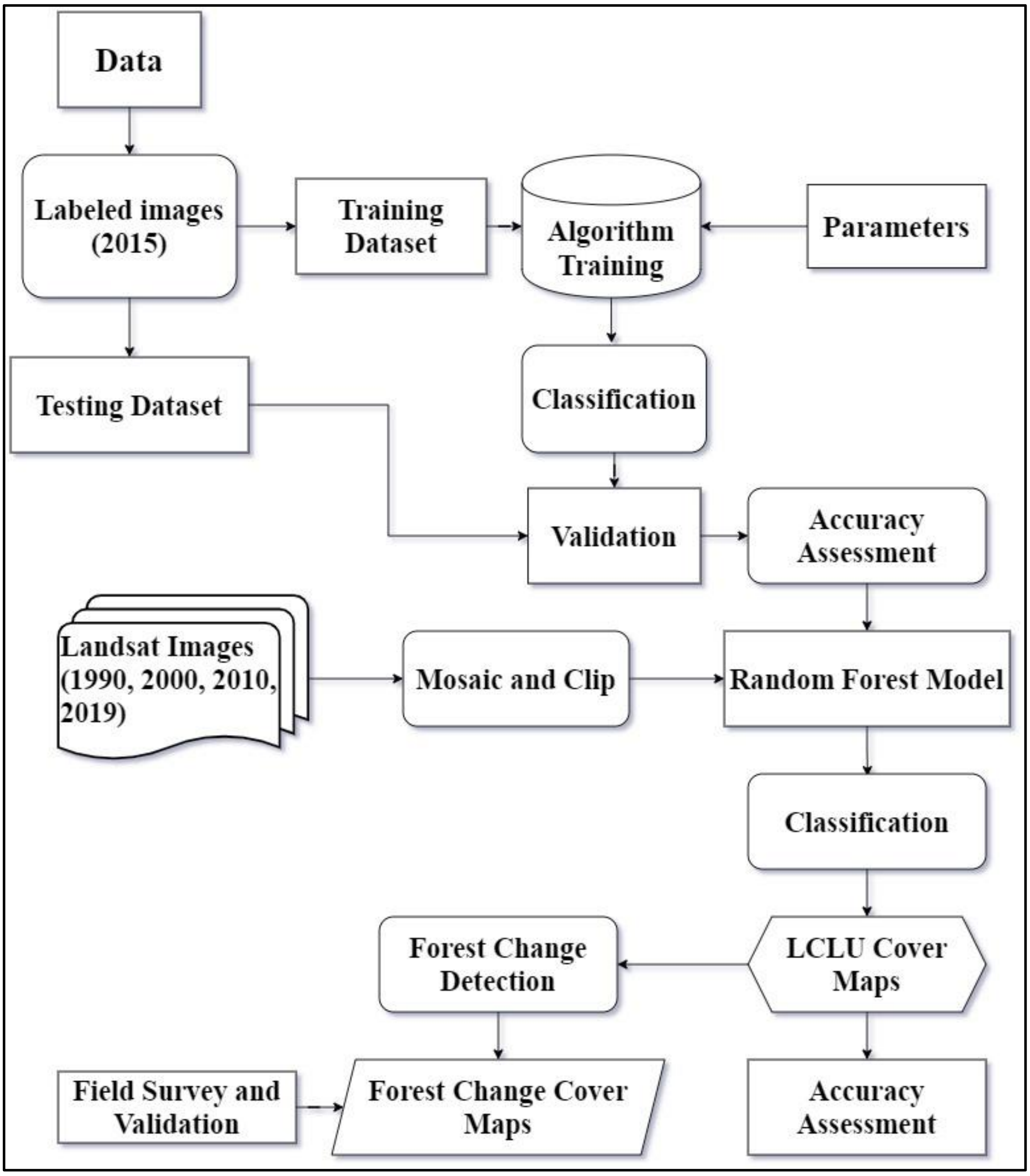


Figure 2.4. Research methodology flow chart.

(http://glovis.usgs.gov/) (Sleeter et al., 2013) for the year 1990, 2000, 2010, and 2019, contemplating the growth cycles of the forest species. Geo-referencing of images was done using image-to-image rectification using GCPs based on maps provided by the forest department of KPK province (Giri, 2016). Extraction of the area of interest with the help of district boundary (Geographical topo-sheet of Pakistan 2003). Labeled images for year 2015 were obtained from forest department Upper Kaghan region, which was used for training and testing and validated machine learning classifiers i.e. (Random Forest, Support vector machine, Perceptron and Classification and regression tree (CART)).

2.3. Image classification model

There were different machine learning algorithms for classification and identification of (forest) land-cover and land-use. We used four machine learning classifiers (Random Forest, Support vector machine, Perceptron and Classification and regression tree) for classification of study area. The model with high accuracy of training and testing was be use for full analysis.

2.3.1 Random Forest (RF)

Random forest is an accurate classifier and always outclass other classifier models based on decision tree algorithms (Çömert et al., 2019). Fast and easily worked model because of its nature to generate bias from single tree decisions (Nguyen et al., 2020). Its classification accuracy will be high when dealing with an area consisting of mountains and hills (Hastie et al., 2009). Using the Random Forest Classifier and making some algorithms changed to uplift the classification accuracy of the model using Python Script (Liaw et al., 2002). Including the additional data in the form of DEM, slope, and aspect the classification accuracy of the Random Forest increased to some noticeable extend.

The RF model had two sensitive parameters: the amount of classification trees to produce (k) and the number of predictor variables used as each node (m) (R. Galiano et al., 2012). 20-fold cross-validation was applied to indicate the amount of decision tree (k=2000) and the number of variables to use at each node (m=3) and maximized the accuracy by minimizing the noises and extra classified classes (Hethcoat et al., 2019). Mean squared error (MSE) is used to convert data branches from each node (m). Intricate branches might lead to time taking the process of training and testing (Gao et al., 2020) (labeled images).

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\mathrm{f}\,i - \mathrm{y}i) \text{ .... Equation 2.1.}$$

* Where:

N is the number of data points

f_i is the value returned by (RF) model

y_i is the actual value for point i

GINI index is used to a defined number of nodes and branches on each node (Lima et al., 2020). Index ensures the prediction likely to occur in each compartment of the decision tree.

$$\mathrm{GINI} = 1 - \sum_{i=1}^{C} fi(1 - fi) \text{ .... Equation 2.2.}$$

* Where:

C is the number of unique labels

f_i is the frequency of label i at a node by (RF) model

Entropy leads to total data in a decision tree (nodes, branches, leaf and root). The pattern of branches should look at random, which is defined by entropy (Hethcoat et al., 2019). Uncertainty was avoidable due to the insertion of the logarithm (log) function in its calculation.

Entropy $= \sum_{i=1}^{C} fi \, log(fi)$ ……………………………… Equation 2.3.

* Where:

C is the number of unique labels
f_i is the frequency of label i at a node by (RF) model

For accurate classification number of nodes, branches, leaf and roots must be in an optimum proportion (Hethcoat et al., 2019). The algorithm must be trained and tested by setting with different GINI and entropy values (sensitivity should be analyzed).

2.3.2 Support vector machine (SVM)

A Support Vector Machine (SVM) is a computational method with a separating hyperplane as its formal definition. The algorithm produces an ideal hyperplane that categorizes fresh samples given labelled training data (supervised learning). This hyperplane is a line that divides a plane into two halves in two-dimensional space, with each class on either side (Meyer et al., 2003).

A non-probabilistic binary linear classifier, an SVM training technique creates a model that assigns new instances to one of two categories (Zhang et al., 2004). The instances in the SVM model were represented as points in space, mapped so that the examples of the different categories were separated by a distinct gap as broad as possible. When data is unlabeled, supervised learning is impossible, and unsupervised learning is the only option (Krogh, 2008). When data is unlabeled, supervised learning is impossible (Hur et al., 2010), hence an unsupervised learning strategy is necessary, in which the data is clustered naturally into groups and then fresh data is mapped to these groups.

2.3.3 Perceptron classifier

Frank Rosenblatt devised the perceptron algorithm at the Cornell Aeronautical Laboratory in 1957, with funding from the US Office of Naval Research (Suykens et al., 1999).

This machine was made to recognize images, with an array of 400 photocells connected to the "neurons" at random. Weights were coded in potentiometers, and electric motors updated the weights during learning (Zhao et al., 2012).

The perceptron is a supervised learning technique for binary classifiers in machine learning. A binary classifier is a function that can determine whether or not a vector of numbers representing an input belongs to a given class (Raudys, 1998). It's a form of linear classifier, or a classification algorithm that uses a linear predictor function to combine a set of weights with the feature vector to create predictions (Hong et al., 1995).

To acquire the optimal split to discriminate observations depending on the dependent variable, rules based on variable values were chosen. The same method is applied to each "child" node when a rule is chosen that splits a node into two (i.e., it is a recursive procedure).

2.3.4 Classification and regression tree (CART)

CART is a non-parametric decision tree learning technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numeric (Chipman et al., 1998).

Splitting comes to an end when CART determines that there is no more gain to be made, or when some pre-determined stopping rules were fulfilled. A terminal node is found at the end of each branch of the tree. Each observation belongs to one and only one terminal node, and each terminal node is defined by a set of rules that were unique to it (Loh, 2011).

2.4 Training and Testing (labeled images)

Labeled image (2015) from provincial forest department Khyber Pakhtunkhwa (KPK) were extracted according to Mansehra region (study area) boundary using toposheet 2003 and converted into 2900 patches of 32*32 pixels in size to train and validate the four-classifier model (Schnitzler et al., 2019) respectively. 75% of data were used to train the classifier model and the rest of the data for validation purposes (Singh et al., 2017)

In environment monitoring and many other subdomains, identifying the physical feature of the earth’s surface (Land cover) as well as how humans exploit the land (Land use) is a difficult task (Duan et al., 2020). Field surveys or satellite image analysis can be used to execute this (Remote Sensing). While field surveys were more accurate and authoritative, they were expensive enterprises that take a long time to update (Wang et al., 2016).

Some labeled images grasp for representation of various land use and land cover classes (Agricultural land, forest, water, snow cover, grasses/shrubs and barren/built-ups). All other classes were masked and only represented one class in each labeled image.

2.4.1 Parameters

Machine learning classifiers needed parameters like elevation, slope, Aspect, Mean Annual temperature, and Mean Annual Rainfall to enhance its classification accuracy (Smith, 2010).

2.4.2 Cover maps

Image classification is defined as a method of image processing that identifies properties in each image based on their spectral signatures, which were assumed to reflect a wavelength function. Once the model (random forest) was trained and obtained desirable accuracy, images were set for (as input) classification. The models generate forest and non-forest (snow, agricultural land, grass/shrubs, barren land, and water) classes from satellite imageries of 1990, 2000, 2010, and 2019 (as output) to analyze and monitor the forest cover change in the study area. The obtained pixel values of forest class and non-forest in the difference image observed which pixels changed from forest class to non-forest, non-forest class to the forest and which pixels retained their values.

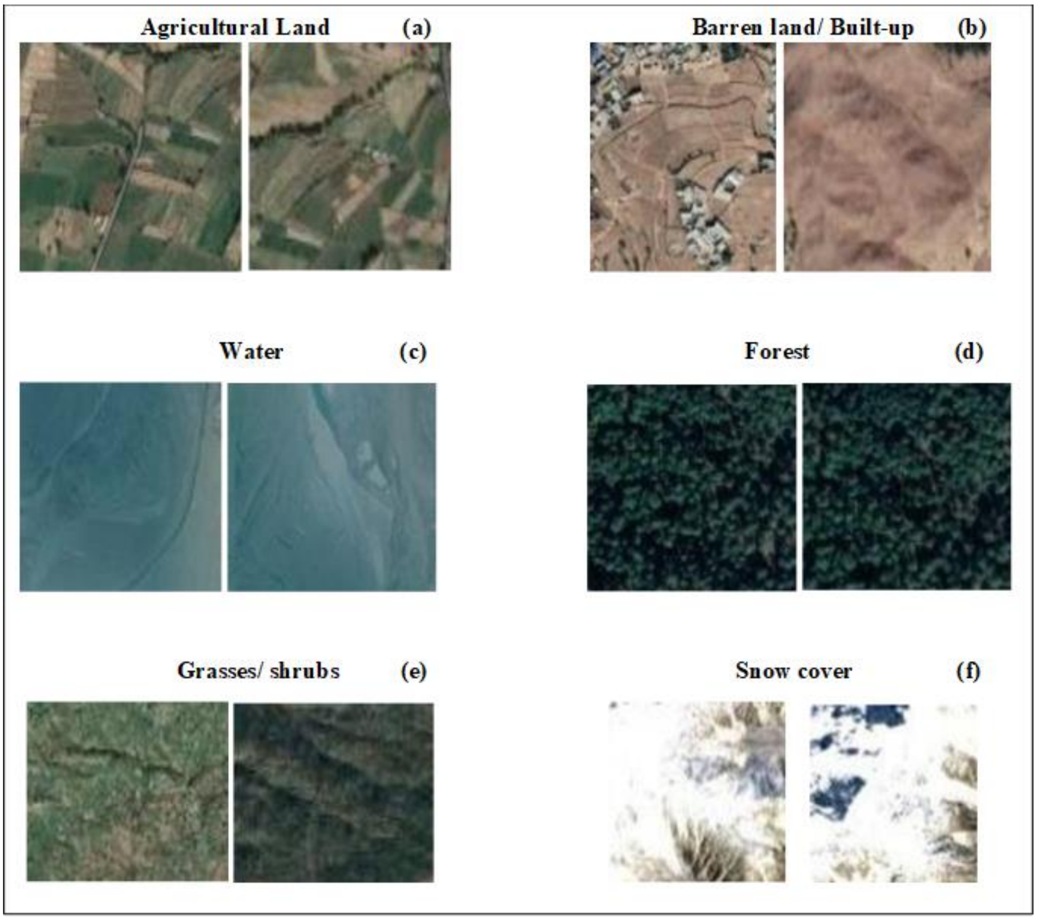


Figure 2.5. Visualization of the different (labeled images) Land-use/land-cover classes.

2.5. Forest cover change detection

Generated cover maps from trained classification models were used to detect the forest changes in the region of interest from 1990 to 2000, 2000 to 2010, 2010 to 2019, the overall change in forest cover was detected for about 30-year 1990 to 2019. For detecting the forest cover change, the image differentiation technique was used, which is fundamentally just the subtraction/minus of the generated cover maps to see which pixels had altered their values (Luppino et al., 2019). The obtained pixel values of forest class and non-forest in the difference image observed which pixels changed from forest class to non-forest, non-forest class to the forest and which pixels retained their values (Thonfeld et al., 2020). Change detection analysis was performed by implementing understandable and straightforward metrics that clearly showed the change in forest class and those areas or pixels that were changed to non-forest classes (Bem et al., 2020).

2.5.1. Change detection metrics

The metrics used for change detection analysis briefly explained below:

I. Forest percentage

Forest % = number of forest class pixels * 100 / number of total pixels ……. Eq 2.4.

II. Percentage forest gain

Forest gain was calculated between consecutive years using the following metric

Forest gain % = number of gain pixels * 100 / number of total pixels ……. Eq 2.5.

* Where the number of gain pixels is the number of pixels that changed from non-forest to forest class between those two years.

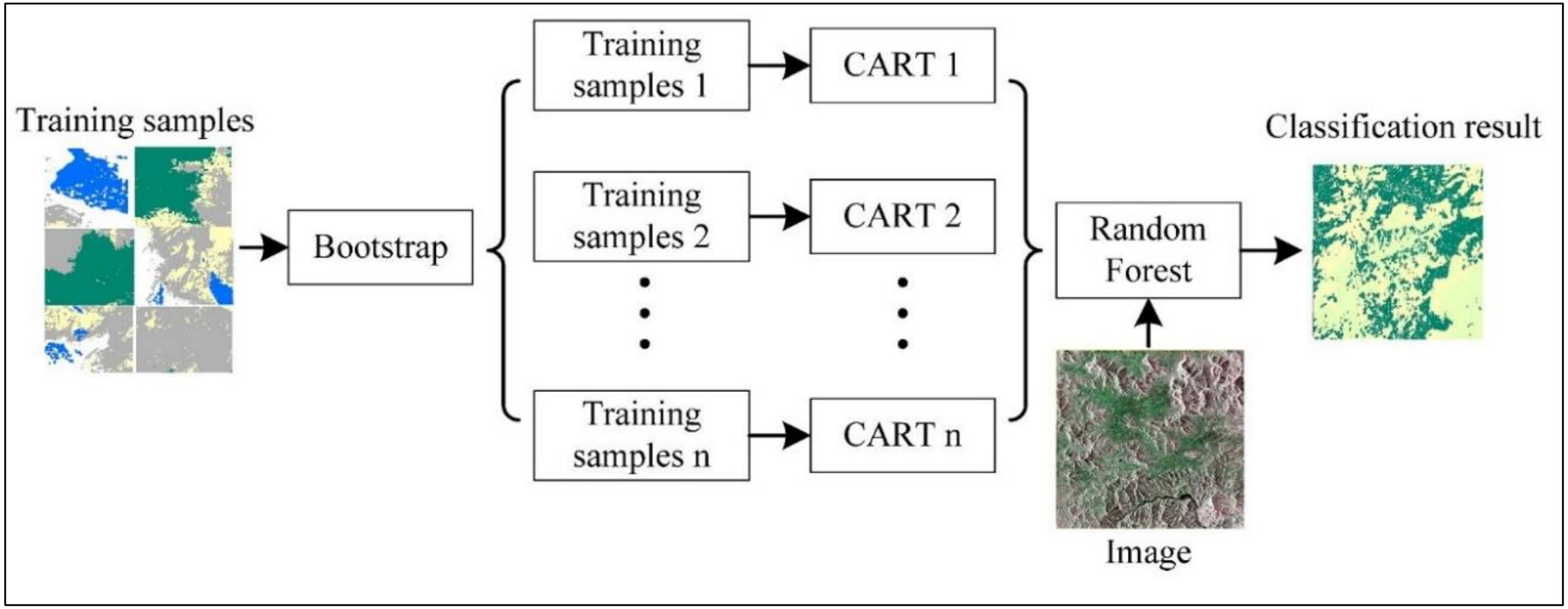


Figure 2.1. Showing the methodology of training samples and the workflow of the random forest model to predict land-cover classes.

III. Percentage forest loss

Forest loss was calculated between consecutive years using the following metric

Forest loss % = number of loss pixels * 100 / number of total pixels …… Eq 2.6.

* Where number of loss pixels is the number of pixels that changed from forest to non-forest class between those two years.

IV. Percentage effective change

Effective forest change % = (% forest gain) + (% forest loss) ………….. Eq 2.7.

* Percentage forest loss always with a negative sign (-).

2.6. Ground survey and validation

Field forest surveys were conducted to ensure the validation and accuracy assessment of forest cover maps generated by the Random Forest model. Four areas were selected randomly and performed field surveys to obtain ground truth figures related to the forest cover area. These areas include Kamalban reserved forest (survey point 1), Kanshian reserved forest near Balakot (survey point 2), Tanglai reserved forest, Kandar valley near Thakot (survey point 3), and Sarin valley in kaghan (survey point 4). Plot method inventory was used in these areas to determine the total area of forest (2019). Compare both results to validate the accuracy of classification and rate of forest cover change for 1990-2019.

Classification accuracy and efficiency of the model were measured based on the overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA) and Kappa coefficient (Rwanga et al., 2017). The value of kappa coefficient should be close to 1, for high and accurate classification (Stehman et al., 2019).

2.7. Forecasting forest cover using exponential smoothing (ETS) algorithm

Time-series data were measured over some time. Forest cover data for the year 1990, 2000, 2010, and 2019 (ten-year intervals) were used to forecast forest cover for the next 40 years (2020-2060), using the exponential triple smoothing (ETS) algorithm (Mahajan et al., 2018).

$$Ft = (\alpha \times At - 1) + (1 - \alpha)\, Ft - 1 \quad \text{.. Equation 2.8.}$$

* Where:

F_t is the forecast forest cover for year t

α is the smoothing constant

A_{t-1} is the previous year's Actual Forest cover

F_{t-1} is the previous year's forecast forest cover

The smoothing constant usually lies between zeros to one. 0.4 value for α was calculated by sensitive measurement of observed values (Sanchez et al., 2019). Forecast forest cover values for the years 2030, 2040, 2050, and 2060 were predicted. Forecasting would help to understand the future scenario of forest stands in the study area (Conceptual visualization) (Sanchez et al.,2019). The interval (10 years) may be slightly predicting forest cover because data were missing between consecutive years (Barrow et al., 2020) (20% missing values would generate possible or near to accurate prediction).

Chapter 3

RESULTS AND DISCUSSIONS

3.1 Comparison of machine learning algorithms

Many scientific organizations working on climate change research, sustainable development, geomorphology, social knowledge management of natural resources, and agricultural land monitoring use detailed land cover maps as a significant input. Furthermore, the capability of land cover monitoring is currently getting expanded by using dense time series, which necessitates efficient, cost-effective categorization systems that had become possible thanks to technological advancements.

Different Algorithms that were discussed in chapter 2 were applied to Mansehra regions. This was to decide which approach, was the best for the segmentation of forest maps on our regions so that we can apply on the best-performing methods for complete analysis. The classifiers (Random Forest, Support vector machine, CART and perceptron) were applied one by one and their confusion matrices and the accuracy were calculated. Table 3.1 shows the comparison, and Figure 3.1 shows the training and test accuracies.

Random Forest classifier clearly outperformed the other statistical machine learning methods and was used for the complete analysis of classification (land-cover/ land-use) and forest cover change detection over Mansehra Region. Distribution patterns might depend on the changing elevation of the region. Snowfall mostly occurred in high altitudes areas like Naran, Kaghan, Lalosar, and Babusar.

Table 3.1. Comparison of classification models based on both training and testing Accuracies.

Classifier	Training Accuracy %	Testing Accuracy %
Random Forest (RF)	89.47	87.58
Support Vector Machine (SVM)	66.33	46.81
Classification And Regression Tree (CART)	73.24	78.48
Perceptron	49.95	51.71

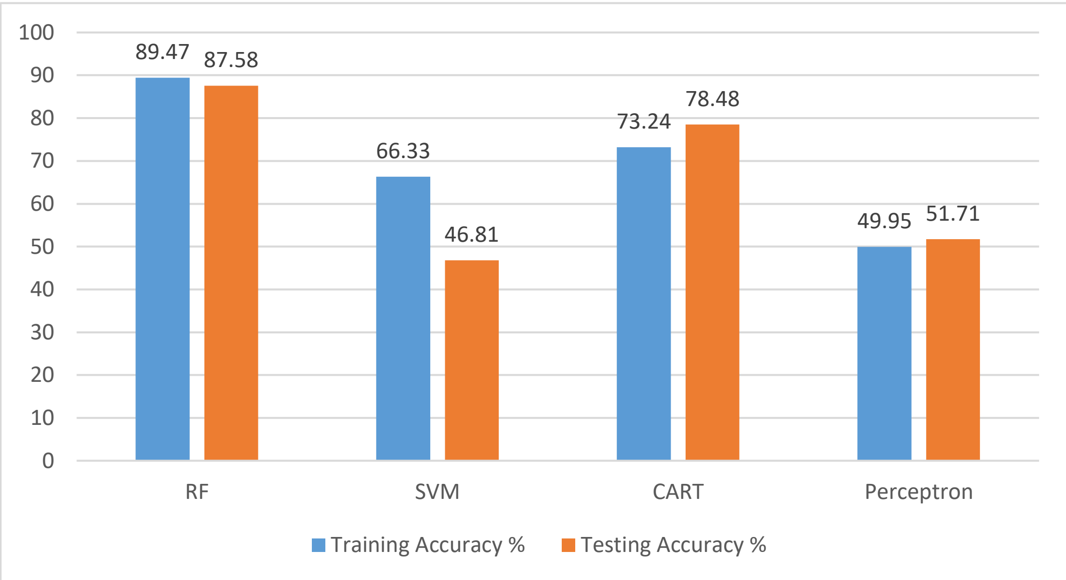


Figure 3.1. Comparison of classification models based on both training and testing accuracies.

Training accuracy percentage was calculated for RF, SVM, CART and Perceptron as 89.47, 66.33, 73.24 and 49.95, respectively. The testing accuracy percentage calculated for RF was 87.58, SVM was 46.81, CART was 78.48 and Perceptron was 51.71.

3.2. Random forest classification (general landcover distribution)

Four land-cover maps consist of six land-cover classes were primed for the study area (Mansehra region). The previous and current distribution during three decades was given in table 3. According to land-cover maps generated for the year 2019, in general, 1161.9660 km^2 of agricultural land, 26.7152 km^2 of water, 473.5345 km^2 of snow cover, 401.5893 km^2 of grasses/shrubs, 1461.0676 km^2 of forest and 93.0748 km^2 of bare soil/rocks. Distribution patterns might depend on the changing elevation of the region. Snowfall mostly occurred in high altitudes areas like Naran, Kaghan, Lalosar, and Babusar (3000-4500 m).

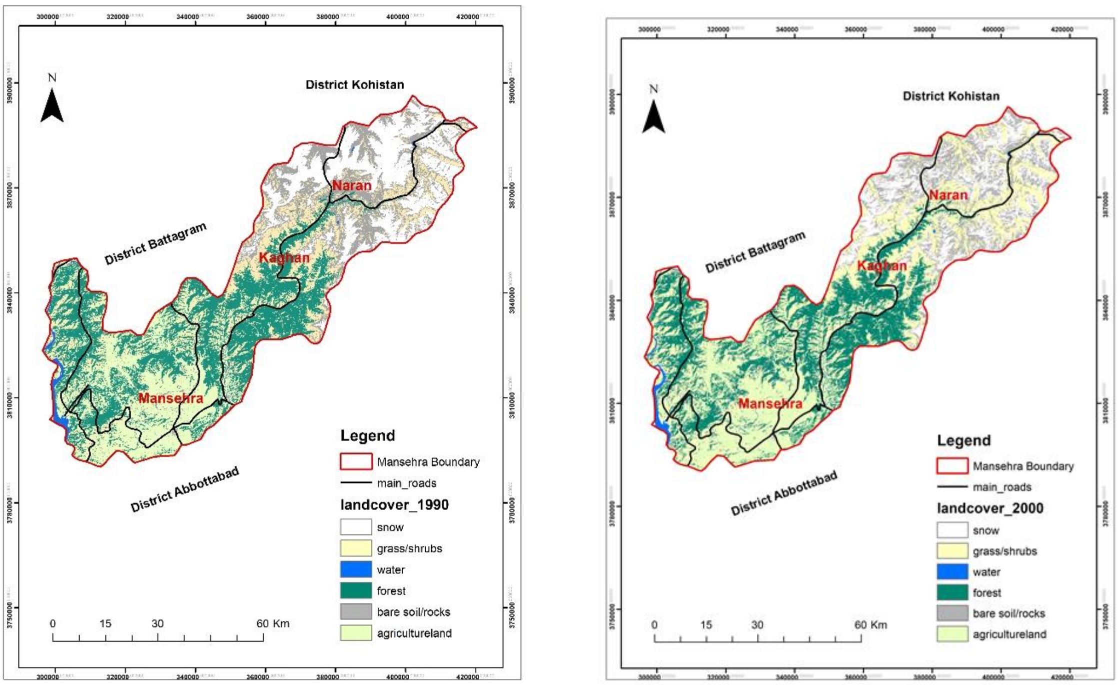
Cities (Mansehra, Ogiri, thakot and kalakot) lie at altitudes of (2500m), mainly comprised of built-ups and road networks. Agricultural practices were mainly carried out in slightly plan areas like Baffa, Nokot, Parehna, Gali Badral, Takia Sharif, and only the primary livelihood resource.

In year 1990, 2000 and 2010 the land-cover distribution was different than the current scenario of the study area. 20.08%, 23.15%, 24.72% of agricultural land; 0.58%, 0.61%, 0.57% of water; 14.90%, 11.86%, 13.01% of snow/glaciers; 13.10%, 14.77%, 17.14% of grasses/shrubs; 36.32%, 31.07%, 27.43% of forest and 12.70%, 16.04%, 12.02% of bare soil/rocks were distributed, respectively.

Table 3.2. Temporal Land-cover/ Land use statistics of Mansehra region.

Landcover	1990	2000	2010	2019
Agricultural land	858.5477	989.4169	1056.6219	1161.9660
Water	24.2664	25.7289	24.5491	26.7152
Snow cover	636.7824	506.7052	556.1338	473.5345
Grasses/Shrubs	559.9571	631.2745	732.5117	401.0748
Forest	1495.6898	1279.4515	1129.5112	1416.0676
Bare soil/rocks	542.7057	685.3571	513.7358	693.0748

*Areas were calculated in terms of square kilometers (Km^2).



(a) (b)

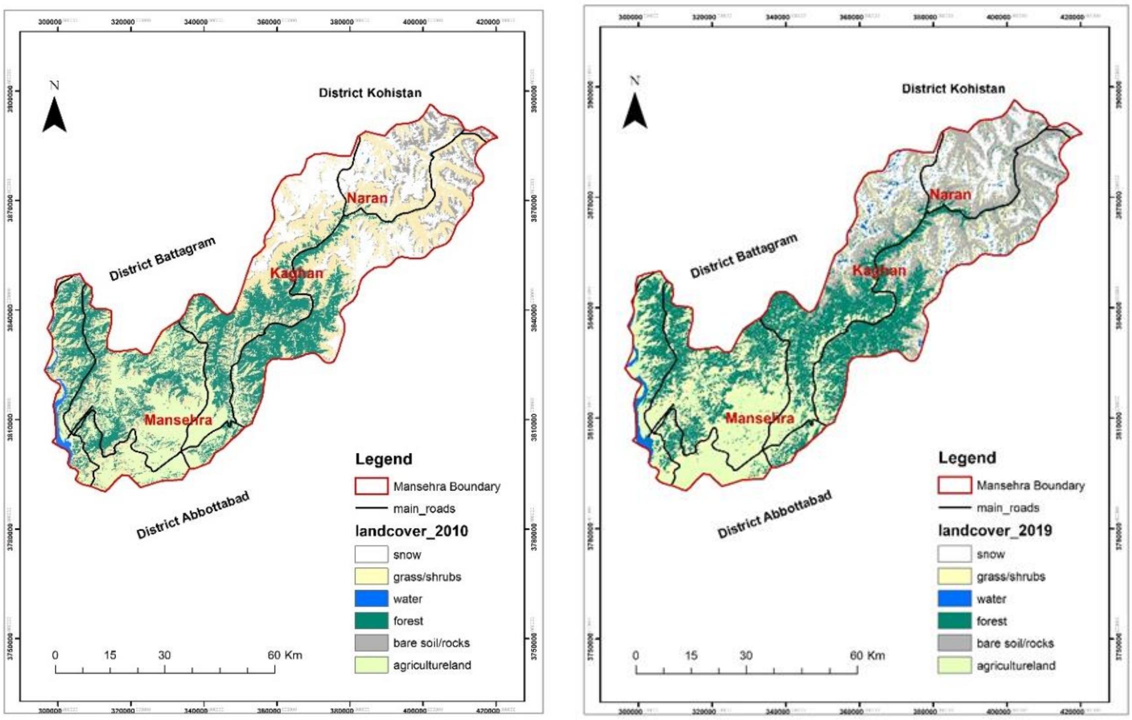
Figure 3.2. Showing the pattern of land-cover classification generated using Random Forest classifier for Mansehra region for year (a) 1990 and (b) 2000.

Forested and agricultural lands were slightly shifted to built-up areas due to the development and changes in the lifestyle of locals. Mansehra city almost lost its agricultural lands, and most people deforested lands to construct houses other infrastructure activities. Areas like Potha (2%), Lohar-banda (0.7%), Faizabad (0.3%), and Gullabad (0.1%) of agricultural practices were carried out in recent years.

The results show that the low altitude areas belong to built-up and agricultural lands. The snow cover mainly occurred in Kaghan, Naran, Shogran, lorasar, Babusar, Sarin valley, and Kaleemabad. Forested areas in 1990 were distributed in the whole region, but it was degraded in the years 2000 and 2010. In 2019 the forest was reforested and preserved by the provincial government and local community efforts by the Billion tree tsunami project, carried out in 2013-2014.

Maps also depict that the agricultural and forested areas were slightly shifted towards high altitude regions. However, water reservoirs were much the same throughout the decades and significant sources of drinking, irrigation, and recharge the groundwater table. In collaboration with the local government, World Bank and Asian development banks take majors to conserve natural resources and biodiversity in the locality by introducing technical skills development programs in 2016.

Urban sprawl was mainly one of the problems in shifting or changing agricultural and forested lands into built-ups. Grasses and shrubs were distributed in the whole area. Barren soil/rocks were increased in 2010 and 2019 due to soil erosion and land sliding in higher study regions. In 2000 and 2010, agricultural practices were increased by replacing forested and bare soil areas.



(a) (b)

Figure 3.3. Showing the pattern of land-cover classification generated using Random Forest classifier for Mansehra region for year (a) 2010 and (b)2019.

3.3. Forest cover change detection

Forest cover changes were detected for 30 years in the study area. The forest's total area was 1495.9571, 1279.4715, 1129.5112, and 1461.0676 Km^2 in 1990, 2000, 2010, and 2019. 68% of the forest in the Mansehra region was consists of coniferous forests, and the rest constitutes a mixed forest. There was a high dispersion of forest classes in the locality, mainly even and uneven forest types. In the year 1990 and 2019 forest cover were detected maximum. Dominating species in areas of Kaghan (37%), Naran (32%), Shogran Valley (21%), and kanshian valley (11%) are Cedrus deodar (deodar), Pinus roxburghii (chir pine), Pinus wallichina (blue pine) and Abies pindrow (fir).

In city areas like Mansehra city, Faizabad, Lahar-banda, Gullabad and Potha, the forest cover area was insignificant. Areas of Thakot (4%), Kalakot (7%), Kanishan valley (11%), and Kaleemabad (3%) were detected as an afforested region of study area under the project of billion tree tsunami.

3.3.1. Forest cover change detection (1990-2000)

The present study identified and temporal forest cover change detection over 30 years (1990-2019) for the Mansehra region's seven tehsils. Forest cover in 1990 was more as compared to 2000, 2010 and 2019. Forest change detection was first carried out for the year 2000 (as compared to 1990). Forest cover percentage in the year 1990 was calculated as 36.32% (1495.6898 km^2), and forest cover percentage in the year 2000 was detected as 31.07%. The result generated by change detection metrics described percentage forest gain in 2000 was 0.02%, percentage forest loss was 5.25%, and percentage significant forest change was -5.23% (negative sign shows deforestation and forest degradation).

The green color in the map shows the forest retains in ten years, the blue color displays the forest gain (afforestation), and the red color shows the removal or forest loss (deforestation). In areas like Kalakot, Wazirabad, Gullabad, Azamgarah, Jabori, Kiwai, Mahandri, and Behari village, forest cover deteriorated (high percentage of forest loss) in a period of 10 years (1990-2000). Mostly forest/wood was felled down for fuelwood and construction purposes by locals.

3.3.2. Forest cover change detection (2000-2010)

Forest change detection for the year 2010 compared to year 2000 was carried out and depicted total forest percentage as 27.43%, percentage forest gain was 0.007%, and percentage forest loss was 3.64%, and percentage effective forest change -3.63%. Results also described that the percentage of significant forest change was -8.89% in 2010 compared to 1990. The forest cover degraded in 2010 in the form of agricultural lands, baren soils, and rangelands. Another reason for deforestation was a large flood that occurred in 2010 (Khan et al., 2020). Soil erosion and land sliding also contributed to the degradation of woodland in the Mansehra region that year (Masood et al., 2011).

The flood that occurred in 2010 (Masood et al., 2011) would raise the surface water flow of Kunhar river, and extreme runoff from its small and large watershed tributaries affected forest cover of high elevated areas of upper Kaghan and Naran, which include valleys like Kholian, Paras, Manila, Katha, Domel, Manhandri, and Jared.

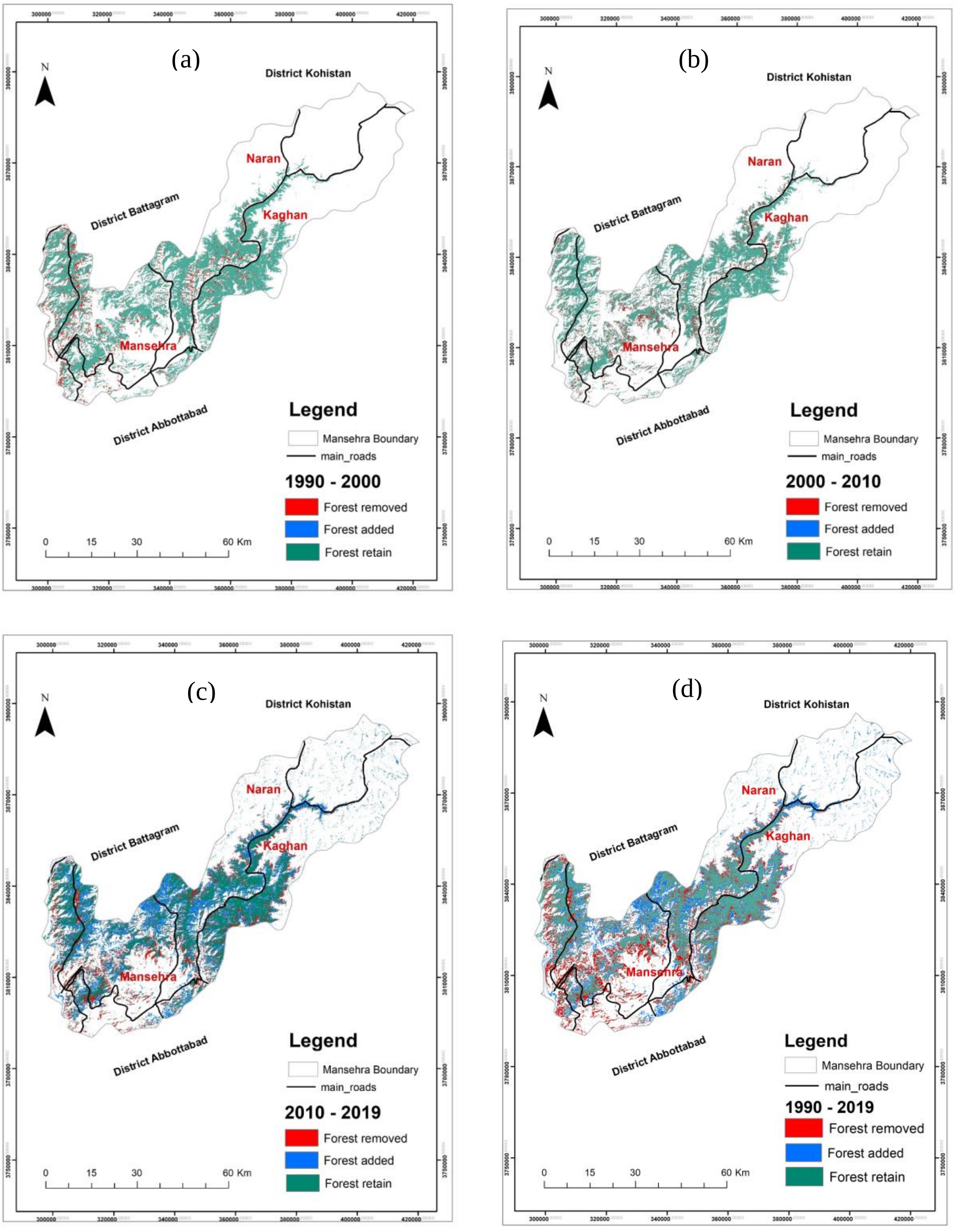


Figure 3.4. Showing the Forest cover change over ten years (a) 1990-2000; (b) 2000-2010; (c) 2010-2019 and (d) 1990-2019.

3.3.3. Forest cover change detection (2010-2019)

Forest conservation is very keen on sustainable resource management. In the year 2013-14 local government and forest departments initiated the "Billion tree tsunami project". Forest change detection metrics calculated total percentage forest as 35.49% in 2019, percentage forest gain was 8.07%, percentage forest loss was 4.14%, and percentage significant change was +4.93% (positive sign indicates afforestation in the region). The forest cover in 2019 was increased compared to 2000 and 2010 by 4.42% and 8.06%.

Forest cover loss occurred in areas near developed cities due to change in livelihood. Forest loss was visible in Roh, Phurzen, Katha, Dobani, Basian, and Shohal. Blue color shows the afforestation or forest gain carried out from 2010 to 2019. An increase in forest cover area was only possible due to "billion tree tsunami projects". In Sherpur, Nokot, Shedore, Hawa Gali, Sharkah, Khohian, Banda Gisach, Ththa, Ghanain, Shungi, Sher Garh, Mohar Khurob, Bahali, Dogah, and Shogran valley, overall forest cover increased about 37%.

Large-scale plantation of forest species in upper Kaghan, Naran, and Behari valley would be fruitful in the coming future. The government also had shown fervid interest in the conservation of forest resources by extending "Billion tree tsunami projects" at the national level (in all provinces, Islamabad, AJK, and Gilgit Baltistan regions) and named as "Ten million tree tsunami projects". Interestingly the current scenario of forest cover in the Mansehra region was the same as that in 1990. Therefore, efforts by the local community must be needed to overcome the forest loss of approximately one percent.

3.3.4. Over-all forest cover change detection (1990-2019)

Overall, forest change detection was carried out by comparing 1990 (first year) and 2019 (second year). The total forest percentage in 1990 was 36.32%, and in 2019 it was 35.49%. Percentage forest gain in 2019 was 9.75%, percentage forest loss was 10.56%, and percentage significant forest change is -0.83%, which was a remarkable achievement related to conservation and afforestation efforts by the provincial government and local communities under the umbrella of "Billion tree tsunami projects." Interestingly the current scenario of forest cover in the Mansehra region was the same as that in 1990. Therefore, efforts by the local community must be needed to overcome the forest loss of approximately 1%.

Some areas were identified and observed the effective forest change to ensure the change detection over some time. In few areas (Biari, Batang and Behri), percentage effective change may be a surprise and very effective, which means the afforestation in some areas generate overall forest gain percentage in past decades. Areas like Nokot, Kiwai and Ghanian regions forest effective change percentages were calculated as -15.38, -17.01 and -2.36, respectively.

In the following figures, the first two images show forest/non-forest binary maps that our model generated for the regions. Green is forest and black is non-forest. The third image is the change image where red means forest loss, blue is forest gain and black is no change (same class sustains). Forest cover change in areas like Wazirabad, Lalakot, Nawabkot, Usman town, Municipal garden town and kunar valley was detected low because most of these retained the total forest cover area in year 2019 as well.

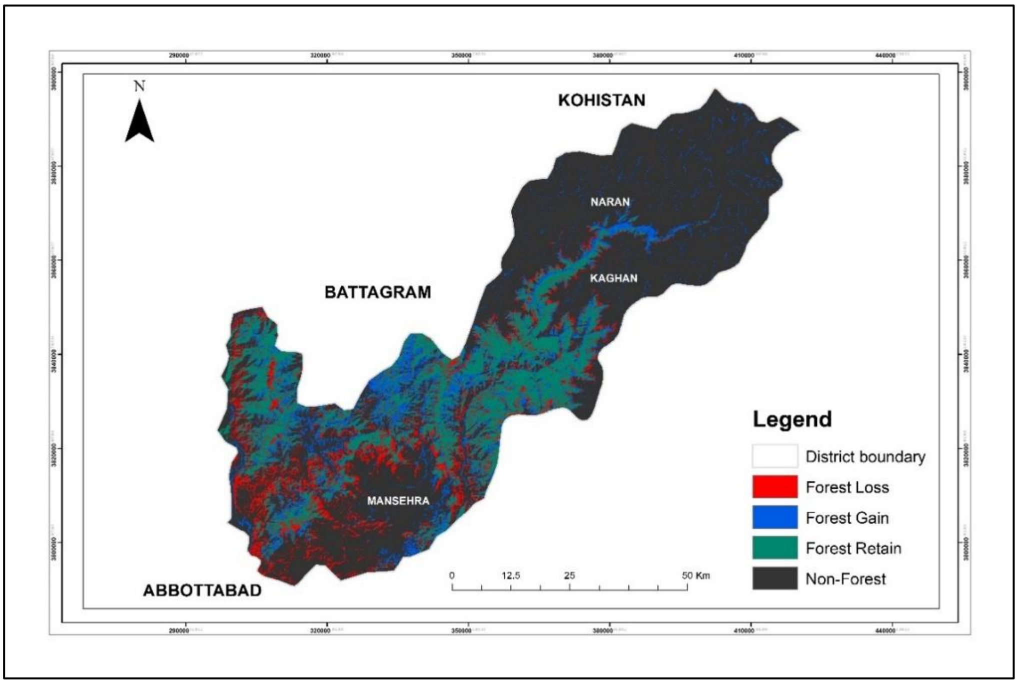


Figure 3.5. Showing the Forest cover change over 30 years (1990 to 2019) in the study area.

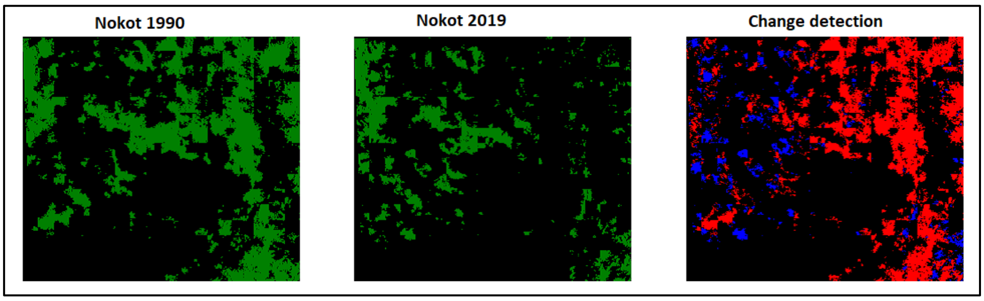


Figure 3.6. Showing Nokot region forest change detection 1990-2019.

Table 3.3. Showing confusion matric results for forest cover change in Nokot Region.

First Year Forest Percentage (1990)	27.37
Second Year Forest Percentage (2019)	11.99
Percentage Forest Gain	3.77
Percentage Forest Loss	19.15
Percentage Effective Change	-15.38

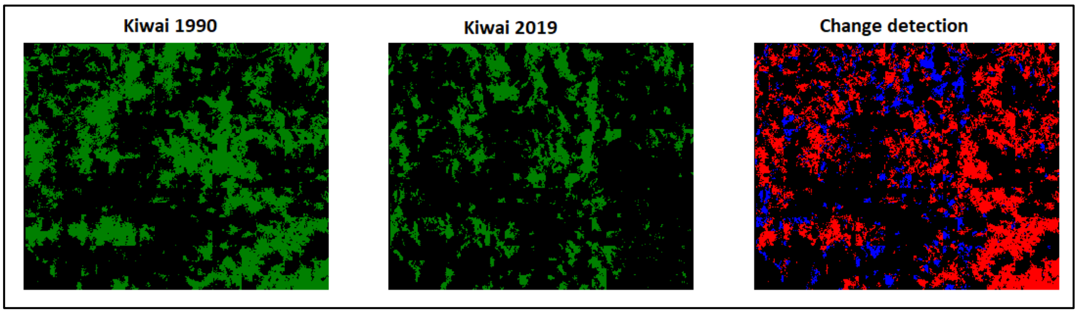


Figure 3.7. Showing Kiwai region forest change detection 1990-2019.

Table 3.4. Showing confusion matric results for forest cover change in Kiwai Region.

First Year Forest Percentage (1990)	34.84
Second Year Forest Percentage (2019)	17.84
Percentage Forest Gain	5.77
Percentage Forest Loss	22.77
Percentage Effective Change	-17.01

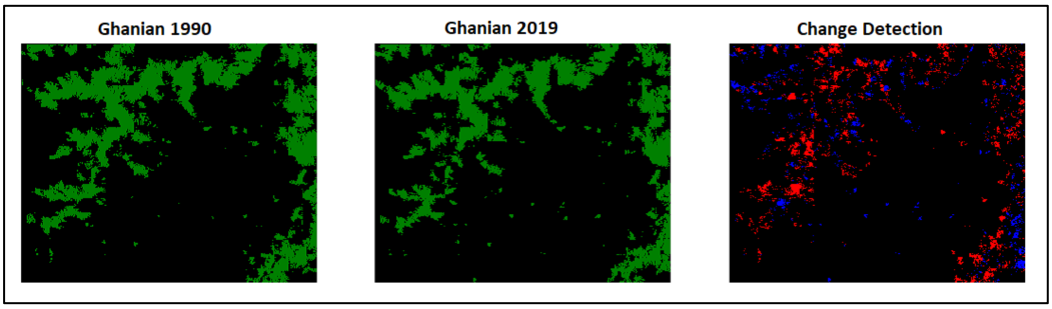


Figure 3.8. Showing Ghanian region forest change detection 1990-2019.

Table 3.5. Showing confusion matric results for forest cover change in Ghanian Region.

First Year Forest Percentage (1990)	20.68
Second Year Forest Percentage (2019)	18.32
Percentage Forest Gain	1.97
Percentage Forest Loss	4.33
Percentage Effective Change	-2.36

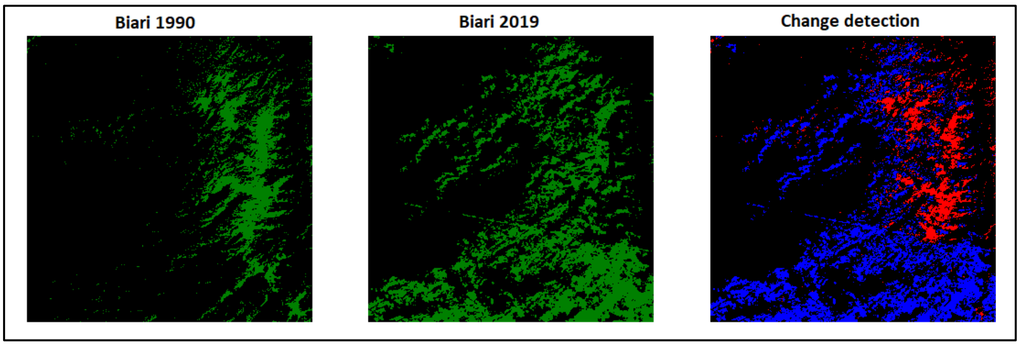


Figure 3.9. Showing Biari region forest change detection 1990-2019.

Table 3.6. Showing confusion matric results for forest cover change in Biari Region.

First Year Forest Percentage (1990)	10.80
Second Year Forest Percentage (2019)	23.15
Percentage Forest Gain	17.15
Percentage Forest Loss	4.80
Percentage Effective Change	+12.35

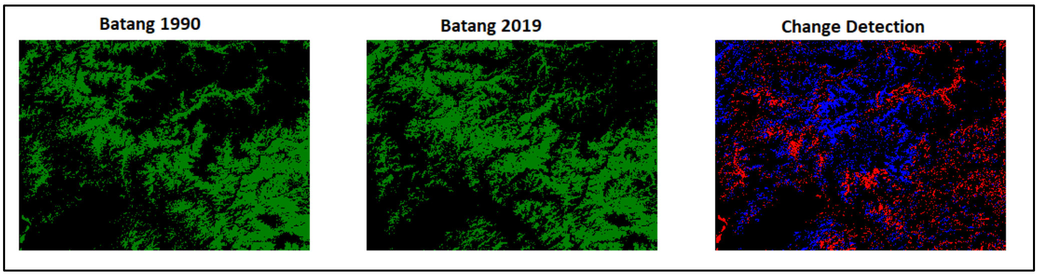


Figure 3.10. Showing Batang region forest change detection 1990-2019.

Table 3.7. Showing confusion matric results for forest cover change in Batang Region.

First Year Forest Percentage (1990)	32.34
Second Year Forest Percentage (2019)	35.50
Percentage Forest Gain	9.90
Percentage Forest Loss	7.20
Percentage Effective Change	+2.70

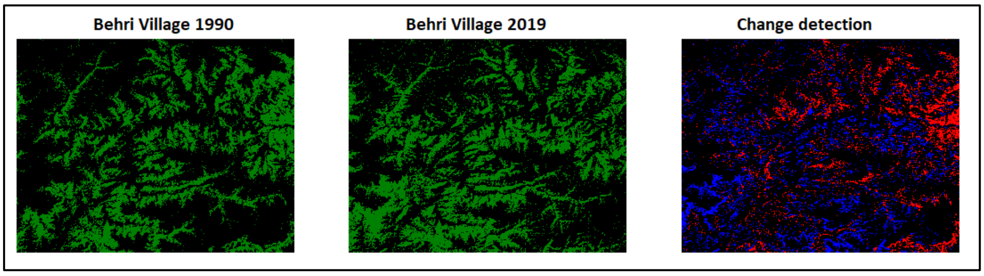


Figure 3.11. Showing Behri region forest change detection 1990-2019.

Table 3.8. Showing confusion matric results for forest cover change in Behri Region.

First Year Forest Percentage (1990)	27.79
Second Year Forest Percentage (2019)	30.42
Percentage Forest Gain	10.83
Percentage Forest Loss	7.46
Percentage Effective Change	+3.37

3.4. Model accuracy and validation

Accuracy assessment of (Random Forest classifier) model classification was performed to ensure the results generated were up to the standards (Significant accuracy). By applying the confusion metrics technique, classification accuracy was systematically determined for the years 1990, 2000, 2010, and 2019, and summarized in Table 3.9. The overall accuracy (OA) percentage of the Random Forest classifier was 89.4%, which was consider significant. However, some LCLU classes' accuracy predicted slightly low because of the season's alteration (annual) in our area of interest. Changing elevation abruptly, also affected the overall accuracy of the classifier. Using a machine learning approach, classification accuracy may be altered and improved compared to traditional classification techniques.

3.4.1. Ground survey for validation of forest cover (2019)

A ground survey was performed at four survey points to collect ground truth data of forest cover in the study area. Compared both ground reality and results generated by Random Forest classifier for the year 2019 at sampled locations.

3.4.1.1. Survey point 1 (Kamalban reserved forest) and point 2 (Kanshian forest)

Filed surveys were carried out in few localities because it was impossible to conduct a survey in the whole region due to lack of funds. Kamalban reserved forest is located in the upper Kaghan (latitude: 34.6829116; longitude: 73.1061214) region.

Table 3.9. Summary of systematic accuracy assessment results.

LCLU	**1990**		**2000**		**2010**		**2019**	
	PA%	**UA%**	**PA%**	**UA%**	**PA%**	**UA%**	**PA%**	**UA%**
Agricultural	78.2	93.9	90.3	87.9	86.2	90.3	84.6	86.7
Water	84.3	88.0	86.4	94.3	86.1	86.1	95.9	94.0
Snow cover	74.9	92.0	84.1	84.1	93.5	82.7	94.1	96.0
Grasses/Shrubs	76.3	88.4	87.2	81.7	83.3	80.6	82.6	86.3
Forest	92.8	83.0	96.6	96.6	94.7	95.8	89.5	90.0
Bare soil/rocks	82.3	92.9	93.5	85.7	88.9	91.7	93.3	84.6
Overall Accuracy	**86.4%**		**89.8%**		**87.5%**		**89.3%**	
Kappa coefficient %	**0.85**		**0.88**		**0.86**		**0.87**	

*PA: Producer's accuracy, UA: User's accuracy.

25 GPS points were used to cover the area of 0.7976 Km^2 (forested), considered ground truth. The model predicted areas of forest in the same locality was 0.7809 Km^2. Survey point 2 was located near Balakot (latitude: 34.538173; longitude: 73.435749). A total of 24 GPS points were taken to measure the forest cover of Kanshian forest and forest cover observed was 0.7108 Km^2, while the model predicted for the same areas was 0.7061 Km^2.

3.4.1.2. Survey point 3 (Tanglai reserved forest) and point 4 (Kandar valley)

Forest cover observed at survey point 3, Tanglai reserved forest (latitude: 34.4771962; longitude: 73.41391952) was 0.4843 Km^2. Classifier model (RF) predicted forest cover for the same area was about 0.4791 Km^2, almost same as ground truth findings. Survey point 4 was Kandar Valley near Thakot (latitude: 34.7311814; longitude: 73.9214743), 26 GPS points were taken, and 0.6598 Km^2 of forest cover was observed in the particular area. The model predicted forest cover area was about 0.6511 Km^2, which shows its supreme accuracy and efficiency.

Steep slopes and abrupt changing elevation of the region might cause some forest cover change detection and classification process challenges. Validation of results may not carry out over the whole area due to low budget. The full field survey may consume time, labor, and capital at a large scale; therefore, validation of results was restricted to possible areas only. Forest regeneration takes a time of about 5 years to grow to the full extent, that's why few areas had forest species, but their growth would not fit that of "Forest Definition", was not detected as forest by the model classifier (forest gain might be high in some areas of the region). Random Forest was most suitable to minimize these effects according to predicted results.

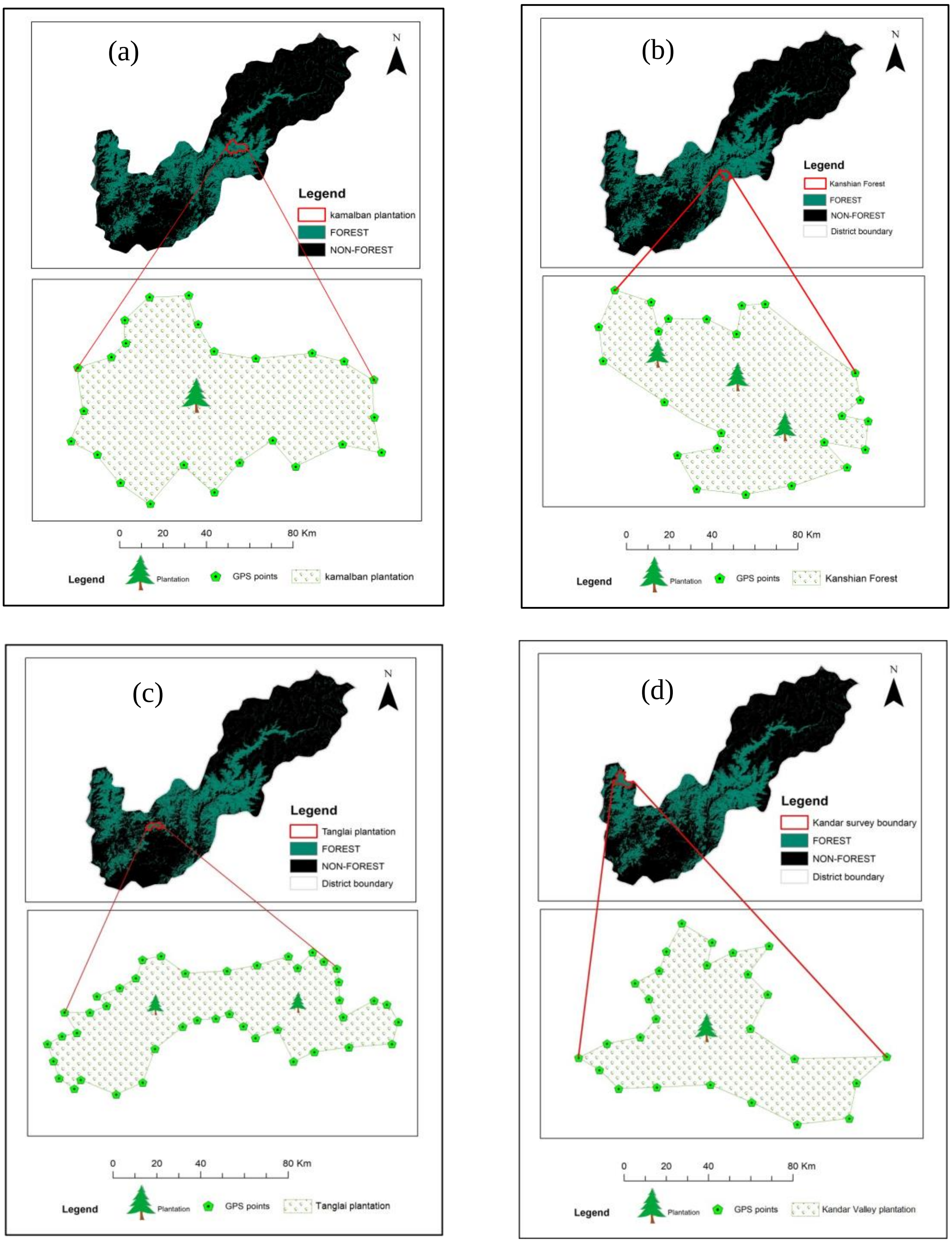


Figure 3.12. Comparison of forest cover observed by ground survey and forest cover area generated by model (RF), Showing actual forest cover and predicted forest cover (a) kamalban reserved forest; (b) Kanshian forest; (c) Tanglai reserved forest and (d) Kandar valley.

3.5. Forecasting forest cover

Time-series data were measured over some time. Forest cover data for the year 1990, 2000, 2010, and 2019 (ten-year intervals) were used to forecast forest cover for the next 40 years (2020-2060), using the exponential triple smoothing (ETS) algorithm.

The result shows a decreasing trend in forest cover for the future (40 years, 2020-2060). The prediction was not satisfied because the impact of missing data might change the expected results. Forest cover for the years 2030, 2040, 2050, and 2060 forecasted as 1348.4305 Km^2, 1338.8774 Km^2, 1319.3243 Km^{2,} and 1304.7712 Km^{2,} respectively.

The decreasing trend in forest cover might vary due to missing data. If intervals were set on a seasonal basis (2-5 years), the prediction would be reversed (increase in forest cover area). Forest cover data for each year was not available. However, this was overwhelming because the total area forecast for the year 2060 was even higher than forest in 2000 (1279.4515 Km^2) and 2010 (1129.5112 Km^2).

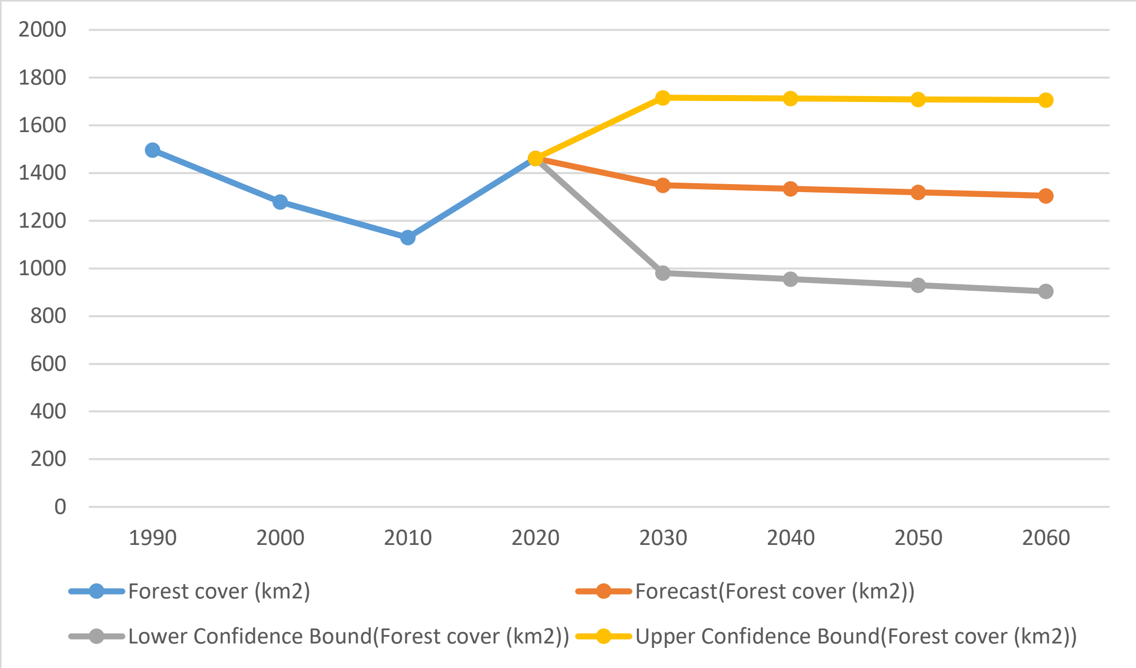


Figure 3.13. Showing time series Forecasting of Forest cover over 40 years (2020-2060) in the study area.

Chapter 4

CONCLUSION AND RECOMMENDATIONS

4.1. Conclusions

The study sets the platform for automatic and real-time monitoring and analyzing natural resources (forest cover) in Pakistan and provides afforestation and deforestation statistics in the Mansehra region (Six Tehsils), for 30 years using publicly available datasets and methods (open source). Random Forest classifier clearly outperformed the other statistical machine learning methods and was used for the complete analysis of classification (land-cover/ land-use) and forest cover change detection over Mansehra Region. Training accuracy percentage was calculated for RF, SVM, CART and Perceptron as 89.47, 66.33, 73.24 and 49.95, respectively. The testing accuracy percentage calculated for RF was 87.58, SVM was 46.81, CART was 78.48 and Perceptron was 51.71.

Forest cover changes were detected for 30 years in the study area. The forest's total area was 1495.9571, 1279.4715, 1129.5112, and 1461.0676 Km^2 in 1990, 2000, 2010, and 2019. 68% of the forest in the Mansehra region was consists of coniferous forests, and the rest constitutes a mixed forest. There was a high dispersion of forest classes in the locality, mainly even and uneven forest types. Dominating species in areas of Kaghan (37%), Naran (32%), Shogran Valley (21%), and kanshian valley (11%) were Cedrus deodar (deodar), Pinus roxburghii (chir pine), Pinus wallichina (blue pine) and Abies pindrow (fir). The results pointed out the places with forest gain (afforestation) and forest loss (Deforestation). The study concluded that the significant increase in afforestation (8.07% forest gain) was predicted between the year 2010 and 2019, but the forest gain percentage between 1990 and 2000 and 2000-2010 was negligible (0.02% and 0.007%, respectively). Deforestation (forest loss) rate was about 5.25%, 3.64% and 4.14% between year 1990 and 2000, 2000-2010 and 2010-2019 respectively. According to the current forest cover change detection scenario, the forest loss percentage was slightly higher (-0.83%) than the forest gain percentage in the Mansehra region (2019).

4.2. Recommendations for further research

The provincial government and local communities' efforts in "Billion tree tsunami projects" played an essential role in protecting and conserving forest resources. Mansehra forest department should put effort to overcome forest loss pattern using State-of-Art mechanisms to monitor its forest products and natural resources.

Fruitful and keen analysis of forest cover change detection also concludes that the involvement of locals in the conservation and protection of forest resources and wildlife habitat was critical for sustainable management. In the future, the extension might help to get more cross-temporal data as well so that a more reliable system can be made for continuous monitoring of forest change. Also, drone imagery could be used to monitor forest health from lower heights, which was not possible using satellite imagery.

REFERENCES

1. Almeida, D., Broadbent, E. N., Zambrano, A. M. A., Wilkinson, B. E., Ferreira, M. E., Chazdon, R., . . . Stark, S. C. (2019). Monitoring the structure of forest restoration plantations with a drone-lidar system. *International Journal of Applied Earth Observation and Geoinformation, 79*, 192-198.

2. Almeida, D. R. A. d., Stark, S. C., Chazdon, R., Nelson, B. W., César, R. G., Meli, P., . . . Moreno, V. S. (2019). The effectiveness of lidar remote sensing for monitoring forest cover attributes and landscape restoration. *Forest Ecology and Management, 438*, 34-43.

3. Awan, M. R., Iqbal, Z., Shah, S. M., Jamal, Z., Jan, G., Afzal, M., . . . Gul, A. (2011). Studies on traditional knowledge of economically important plants of Kaghan Valley, Mansehra District, Pakistan. *Journal of Medicinal Plants Research, 5*(16), 3958-3967.

4. Barrow, D., Kourentzes, N., Sandberg, R., & Niklewski, J. (2020). Automatic robust estimation for exponential smoothing: Perspectives from statistics and machine learning. *Expert Systems with Applications, 160*, 113637.

5. Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines *Data mining techniques for the life sciences* (pp. 223-239): Springer.

6. Chipman, H. A., George, E. I., & McCulloch, R. E. (1998). Bayesian CART model search. *Journal of the American Statistical Association, 93*(443), 935-948.

7. Çömert, R., MATCI, D. K., & Avdan, U. (2019). Object Based Burned Area Mapping with Random Forest Algorithm. *International Journal of Engineering and Geosciences, 4*(2), 78-87.

8. de Bem, P. P., de Carvalho Junior, O. A., Fontes Guimarães, R., & Trancoso Gomes, R. A. (2020). Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks. *Remote Sensing, 12*(6), 901.

9. Duan, J., Li, S. E., Guan, Y., Sun, Q., & Cheng, B. (2020). Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data. *IET Intelligent Transport Systems, 14*(5), 297-305.

10. Gao, Y., Skutsch, M., Paneque-Gálvez, J., & Ghilardi, A. (2020). Remote sensing of forest degradation: a review. *Environmental Research Letters, 15*(10), 103001.

11. Giri, C. (2016). Observation and monitoring of mangrove forests using remote sensing: opportunities and challenges: Multidisciplinary Digital Publishing Institute.

12. Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning *The elements of statistical learning* (pp. 485-585): Springer.

13. Hethcoat, M. G., Edwards, D. P., Carreiras, J. M., Bryant, R. G., Franca, F. M., & Quegan, S. (2019). A machine learning approach to map tropical selective logging. *Remote Sensing of Environment, 221*, 569-582.

14. Hong, J. H., Campbell, S., & Yeh, P. (1995). Optical pattern classifier with perceptron learning *Landmark Papers On Photorefractive Nonlinear Optics* (pp. 551-557): World Scientific.

15. Hussain, S. L., Hussain, A., Khan, S., & Khan, S. (2016). Evaluating the impacts of forest fire on the vegetation of sub-tropical Chir pine forest in district Mansehra, Pakistan. *Pure and Applied Biology, 5*(1), 102.

16. Khalid, N. (2020). Prevalence of Avian Influenza Virus in Poultry Population of District Mansehra. *EC Veterinary Science, 5*, 15-20.

17. Khan, K., Ishtiaq, M., Iqbal, Z., Alam, J., Bhatti, K., Shah, A., . . . Mehmood, A. (2018). Effects of edaphic and physiognomic factors on species diversity, distribution and composition in reserved forest of Sathan Gali (Mansehra), Pakistan. *Appl. Ecol. Environ. Res, 16*(2), 1085-1100.

18. Krogh, A. (2008). What are artificial neural networks? *Nature biotechnology, 26*(2), 195-197.

19. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2*(3), 18-22.

20. Lima, T. A., Beuchle, R., Griess, V. C., Verhegghen, A., & Vogt, P. (2020). Spatial patterns of logging-related disturbance events: a multi-scale analysis

on forest management units located in the Brazilian Amazon. *Landscape Ecology, 35*(9), 2083-2100.

21. Loh, W. Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery, 1*(1), 14-23.

22. Luppino, L. T., Bianchi, F. M., Moser, G., & Anfinsen, S. N. (2019). Unsupervised image regression for heterogeneous change detection. *arXiv preprint arXiv:1909.05948*.

23. Mahajan, S., Chen, L.-J., & Tsai, T.-C. (2018). Short-term pm2. 5 forecasting using exponential smoothing method: A comparative analysis. *Sensors, 18*(10), 3223.

24. Meyer, D., Leisch, F., & Hornik, K. (2003). The support vector machine under test. *Neurocomputing, 55*(1-2), 169-186.

25. Naz, F., Haq, I. U., Asghar, S., Shah, A. S., & Rahman, A. (2011). Studies on growth, yield and nutritional composition of different tomato cultivars in Battal valley of district Mansehra, Khyber Pakhtunkhwa, Pakistan. *Sarhad J. Agric, 27*(4), 569-571.

26. Nguyen, L. H., Joshi, D. R., Clay, D. E., & Henebry, G. M. (2020). Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier. *Remote Sensing of Environment, 238,* 111017.

27. Raudys, Š. (1998). Evolution and generalization of a single neurone: I. Single-layer perceptron as seven statistical classifiers. *Neural Networks, 11*(2), 283-296.

28. Rendon-Sanchez, J. F., & de Menezes, L. M. (2019). Structural combination of seasonal exponential smoothing forecasts applied to load forecasting. *European Journal of Operational Research, 275*(3), 916-924.

29. Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 93-104.

30. Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy assessment of land use/land cover classification using remote sensing and GIS. *International Journal of Geosciences, 8*(04), 611.

31. Schnitzler, N., Ross, P.-S., & Gloaguen, E. (2019). Using machine learning to estimate a key missing geochemical variable in mining exploration: Application of the Random Forest algorithm to multi-sensor core logging data. *Journal of Geochemical Exploration, 205*, 106344.

32. Singh, B., Sihag, P., & Singh, K. (2017). Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Modeling Earth Systems and Environment, 3*(3), 999-1004.

33. Sleeter, B. M., Sohl, T. L., Loveland, T. R., Auch, R. F., Acevedo, W., Drummond, M. A., . . . Stehman, S. V. (2013). Land-cover change in the

conterminous United States from 1973 to 2000. *Global Environmental Change, 23*(4), 733-748.

34. Smith, A. (2010). Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm. *Journal of Spatial Science, 55*(1), 69-79.

35. Stehman, S. V., & Foody, G. M. (2019). Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment, 231*, 111199.

36. Suykens, J. A., & Vandewalle, J. (1999). Training multilayer perceptron classifiers based on a modified support vector method. *IEEE transactions on Neural Networks, 10*(4), 907-911.

37. Thonfeld, F., Steinbach, S., Muro, J., & Kirimi, F. (2020). Long-term land use/land cover change assessment of the Kilombero catchment in Tanzania using random forest classification and robust change vector analysis. *Remote Sensing, 12*(7), 1057.

38. Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). *Cnn-rnn: A unified framework for multi-label image classification.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

39. Ye, S., Rogan, J., Zhu, Z., & Eastman, J. R. (2020). A near-real-time approach for monitoring forest disturbance using Landsat time series:

Stochastic continuous change detection. *Remote Sensing of Environment, 252*, 112167.

40. Zhang, L., Zhou, W., & Jiao, L. (2004). Wavelet support vector machine. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 34*(1), 34-39.

41. Zhao, C., Gao, Y., He, J., & Lian, J. (2012). Recognition of driving postures by multiwavelet transform and multilayer perceptron classifier. *Engineering Applications of Artificial Intelligence, 25*(8), 1677-1686.

42. Afify, H. A. (2011). Evaluation of change detection techniques for monitoring land-cover changes: A case study in new Burg El-Arab area. *Alexandria engineering journal, 50*(2), 187-195.

43. Allan, J. R., Venter, O., Maxwell, S., Bertzky, B., Jones, K., Shi, Y., & Watson, J. E. (2017). Recent increases in human pressure and forest loss threaten many Natural World Heritage Sites. *Biological conservation, 206*, 47-55.

44. Amir, M., Liu, X., Ahmad, A., Saeed, S., Mannan, A., & Muneer, M. A. (2018). Patterns of biomass and carbon allocation across chronosequence of chir pine (Pinus roxburghii) forest in Pakistan: inventory-based estimate. *Advances in Meteorology, 2018*.

45. Bajwa, G. A., Shahzad, M. K., & Satti, H. K. (2015). Climate change and its impacts on growth of blue pine (Pinus wallichiana) in Murree Forest division, Pakistan. *Science, Technology and Development, 34*(1), 27-34.

46. Bohn, F. J., & Huth, A. (2017). The importance of forest structure to biodiversity–productivity relationships. *Royal Society open science, 4*(1), 160521.

47. Champion, S. H., Seth, S. K., & Khattak, G. (1965). Forest types of Pakistan. *Forest types of Pakistan.*

48. Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Review ArticleDigital change detection methods in ecosystem monitoring: a review. *International journal of remote sensing, 25*(9), 1565-1596.

49. Coppin, P. R., & Bauer, M. E. (1996). Digital change detection in forest ecosystems with remote sensing imagery. *Remote sensing reviews, 13*(3-4), 207-234.

50. Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., & Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science, 361*(6407), 1108-1111.

51. Dash, M., & Behera, B. (2016). Determinants of household collection of non-timber forest products (NTFPs) and alternative livelihood activities in Similipal Tiger Reserve, India. *Forest Policy and Economics, 73*, 215-228.

52. Dove, M. R. (1992). Foresters' beliefs about farmers: a priority for social science research in social forestry. *Agroforestry Systems*, *17*(1), 13-41.

53. Duguma, L. A., Atela, J., Minang, P. A., Ayana, A. N., Gizachew, B., Nzyoka, J. M., & Bernard, F. (2019). Deforestation and forest degradation as an environmental behavior: unpacking realities shaping community actions. *Land*, *8*(2), 26.

54. Herold, M., Carter, S., Avitabile, V., Espejo, A. B., Jonckheere, I., Lucas, R., . . . Petersen, R. (2019). The role and need for space-based forest biomass-related measurements in environmental management and policy. *Surveys in Geophysics*, *40*(4), 757-778.

55. Jha, C., Dutt, C., & Bawa, K. (2000). Deforestation and land use changes in Western Ghats, India. *Current Science*, 231-238.

56. Katarki, G., Ranmale, H., Bidari, I., & Chickerur, S. (2019). *Estimating Change Detection of Forest Area using Satellite Imagery*. Paper presented at the 2019 International Conference on Data Science and Communication (IconDSC).

57. Khan, S. H., He, X., Porikli, F., & Bennamoun, M. (2017). Forest change detection in incomplete satellite images with deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing, 55*(9), 5407-5423.

58. Khan, Z., Shah, A. H., Farooq, M., Khan, K. R., Hussain, M., Shah, G. M., . . . Rahman, I. U. (2020). Estimation of fuel wood consumption and its negative impact on surrounding vegetation: A case study of Baffa town, Mansehra, Pakistan. *Proceedings of the International Academy of Ecology and Environmental Sciences, 10*(1), 22-31.

59. Kushwaha, S. (1990). Forest-type mapping and change detection from satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 45*(3), 175-181.

60. Lazdinis, M., Angelstam, P., & Pülzl, H. (2019). Towards sustainable forest management in the European Union through polycentric forest governance and an integrated landscape approach. *Landscape Ecology, 34*(7), 1737-1749.

61. Luo, F., Moyle, B. D., Bao, J., & Zhong, Y. (2016). The role of institutions in the production of space for tourism: National Forest Parks in China. *Forest Policy and Economics, 70,* 47-55.

62. MacDicken, K. G. (2015). Global forest resources assessment 2015: what, why and how? *Forest Ecology and Management, 352*, 3-8.

63. Milne, B. T. (1988). Measuring the fractal geometry of landscapes. *Applied Mathematics and Computation, 27*(1), 67-79.

64. Niraula, R. R., Gilani, H., Pokharel, B. K., & Qamer, F. M. (2013). Measuring impacts of community forestry program through repeat photography and satellite remote sensing in the Dolakha district of Nepal. *Journal of environmental management, 126*, 20-29.

65. Panta, M., Kim, K., & Joshi, C. (2008). Temporal mapping of deforestation and forest degradation in Nepal: Applications to forest conservation. *Forest Ecology and Management, 256*(9), 1587-1595.

66. Park, S. W., & Lee, Y. W. (2019). *Detection of forest disaster using high-resolution satellite images with semantic segmentation.* Paper presented at the Image and Signal Processing for Remote Sensing XXV.

67. Pendrill, F., Persson, U. M., Godar, J., & Kastner, T. (2019). Deforestation displaced: trade in forest-risk commodities and the prospects for a global forest transition. *Environmental Research Letters, 14*(5), 055003.

68. Qamer, F. M., Abbas, S., Saleem, R., Shehzad, K., Ali, H., & Gilani, H. (2012). Forest cover change assessment in conflict-affected areas of northwest Pakistan: The case of Swat and Shangla districts. *Journal of Mountain Science, 9*(3), 297-306.

69. Qamer, F. M., Shehzad, K., Abbas, S., Murthy, M., Xi, C., Gilani, H., & Bajracharya, B. (2016). Mapping deforestation and forest degradation patterns in western Himalaya, Pakistan. *Remote Sensing, 8*(5), 385.

70. Qasim, M., Hubacek, K., Termansen, M., & Fleskens, L. (2013). Modelling land use change across elevation gradients in district Swat, Pakistan. *Regional environmental change, 13*(3), 567-581.

71. Saeed, S., Ashraf, M. I., Ahmad, A., & Rahman, Z. (2016). The Bela forest ecosystem of district Jhelum, a potential carbon sink. *Pakistan Journal of Botany, 48*(1), 121-129.

72. Schulze, K., Malek, Ž., & Verburg, P. H. (2019). Towards better mapping of forest management patterns: A global allocation approach. *Forest Ecology and Management, 432*, 776-785.

73. Shahbaz, B., Ali, T., & Suleri, A. Q. (2007). A critical analysis of forest policies of Pakistan: implications for sustainable livelihoods. *Mitigation and Adaptation Strategies for Global Change, 12*(4), 441-453.

74. Sher, H., Al-Yemeni, M., & Sher, H. (2010). Forest resource utilization assessment for economic development of rural community in northern parts of Pakistan. *Journal of Medicinal Plants Research, 4*(12), 1197-1208.

75. Shiddamallayya, N., Rao, R., Doddamani, S., & Venkateshwarlu, G. (2016). A glimpse on forest flora and Indian system of medicine plants of Chitradurga district, Karnataka. *International Journal of Herbal Medicine, 4*(1), 25-33.

76. Sloan, S., & Sayer, J. A. (2015). Forest Resources Assessment of 2015 shows positive global trends but forest loss and degradation persist in poor tropical countries. *Forest Ecology and Management, 352*, 134-145.

77. Wu, X. (2019). Plant Selection in the Environmental Art Design of Forest Landscape. *Ekoloji*, *28*(108), 1277-1281.

78. Zeb, A., Hamann, A., Armstrong, G. W., & Acuna-Castellanos, D. (2019). Identifying local actors of deforestation and forest degradation in the Kalasha valleys of Pakistan. *Forest Policy and Economics, 104*, 56-64.

79. Zhang, Q., Li, Y., Yu, C., Qi, J., Yang, C., Cheng, B., & Liang, S. (2020). Global timber harvest footprints of nations and virtual timber trade flows. *Journal of Cleaner Production, 250*, 119503.

80. Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., . . . Strobl, P. (2019). Benefits of the free and open Landsat data policy. *Remote Sensing of Environment, 224*, 382-385.