Concept for Mapping Carbon footprint with Change in Vegetation Cover and Population in India

Ishma Amin, Meenakshi Agarwal, Stevert Lobo, Rahul Gurnani, Prof. (Mrs.) Priya R.L (*Authors*)
Computer Department
Vivekanand Education Society's Institute of Technology
Mumbai, India
{2016.ishma.amin, 2016.meenakshi.agarwal, 2016.stevert.lobo, 2016.rahul.gurnani, priya.rl } @ves.ac.in

Abstract—In most of the developing countries, the increasing rate of Carbon emissions is considered as a major cause of concern. India is leading in terms of CO2 emissions as compared to other countries. The vegetation cover comprises only 24.39% of the geographic area of India. Metropolitan cities in India are witnessing rapid urbanization. The primary objective of this proposal is to identify the relationship between the increase in carbon emissions and deforestation in metropolitan cities. An additional objective is to predict the amount of afforestation required for each area to cope up with the carbon emissions over the next 25-years. It can be achieved by using statistical models like ARIMA, LSTM and machine learning techniques such as Random Forest. The proposal provides suggestions on optimal places and techniques for sustainable afforestation to the concerned authorities using artificial intelligence.

Keywords—Carbon Footprint, Deforestation, ARIMAX, VAR, Random Forest, LSTM, k-NN imputation, MICE Imputation, time-series synchrony quantification, machine learning.

I. INTRODUCTION

Carbon footprint is used to measure the amount of greenhouse gases mainly, carbon dioxide (CO_2) released into the atmosphere by urban activities. It is normally estimated as tons of CO_2 produced every year, which can be enhanced by huge amounts of CO_2-identical gases, including methane, nitrous oxide, and other greenhouse gases.[3].

The change in demographic profile, urbanization and increase in the number of vehicles, rapid population growth and industrialization over the years, the carbon content in the atmosphere has risen at an exponential rate.

According to the survey of the Global Climate Risk Index (CRI) of 2017, India is 14th on the list of most vulnerable countries. The global ambient air quality database published by the World Health Organization (WHO) in 2018 specifies that urban areas with the highest levels of PM2.5 were situated in India. In AirVisual's 2018 World air quality report, it was mentioned that most polluted urban areas based on PM2.5 estimates on the planet are located in India.[4].

The ecosystem services provided by tree and forest cover of the country become very essential for the existence of any living being. The panel report of India State of Forest Report 2019 states that nation-wide total forest cover is 21.67% of the geographical area, whereas the tree cover is only 2.89% [5,6]. It is also estimated that Soil Organic Carbon (SOC) defines the largest carbon stack in Indian forest areas.

Existing systems relate carbon footprint and deforestation concerning a product or domain-specific industry. This research aims to calculate the net carbon content, then calculate its relationship with forest cover. This relationship will help us to determine the requirements for the future.

This paper is divided into three sections: Section A discusses the survey of this research and section B describes the proposed idea of the system. Goals of the proposed idea for reducing CO_2 emissions are discussed in section C. It also elaborates various components of proposed architecture and methodologies for determining the impact of deforestation on regional carbon footprint levels. Finally, the merits and challenges of the proposed idea are discussed with the future scope of this project.

II. RELATED WORK

The study [7] speaks to a connection between development and qualities and points. By utilizing a negative binomial relapse model, they built the connection between ranchers' ingenuity and those qualities and points which control ranchers' homestead the executive's choices just as another homestead/rancher attributes. Accordingly, the study found that ranchers who are creative win more. The study offers the initial phase in connecting maintainability esteems and development exercises [7].

The paper [8] analyses the carbon emission of the Birla Institute of Technology and Science Pilani (BITS Pilani), using the life cycle analysis (LCA) system. The immediate emanations of college possessed offices, electricity, warmth or steam, and other circuitous discharges are the results of college exercises that are the source for greenhouse gases. Different discharges are displayed in Umberto NXT Universal programming utilizing ISO 14064 guidelines. [8]

The paper [9] uses Deep neural systems that utilize stepwise linear regressions with exponential smoothing regression slopes as trend strength indicators for a given time interval. The outcomes exhibit the estimation of Deep learning ways to deal with time series analysis and show that regression gives helpful highlights to complex interdependencies, offering a straightforward marker of trends with high predictive value [9].

This paper [10] focuses on forest cover dynamics and creating a forecasting model of the Bhanpratappur Forest

Division of Kanker, Chhattisgarh, India. It uses a logistic regression model (LRM) to analyze the changes in forests caused by road constructions and expansion of settlements. This research uses data obtained through Landsat TM satellite imagery for the duration of 1990 to 2000. The model was assessed by comparing the model-predicted forest cover with the actual forest cover for 2010 [10].

This paper is intended to anticipate the influence of variables such as electricity, coal, etc. in the growth of carbon emissions. The data was obtained from the Alcohol Industry to train and test [11].

The following TABLE I depicts the comparison of various existing systems.

TABLE I.   COMPARISON BETWEEN EXISTING SYSTEMS

Existing System	Proposed System
Negative binomial regression model generating relationship.[7]	The ARIMAX model is used for forecasting and formulating equations.
Deep neural networks for trend change analysis [10].	LSTM based neural networks for trend change analysis are applied.
Logistic regression model [4] and Support Vector Machine [11] for prediction	Time Series analysis is used for prediction

III. PROPOSED IDEA

The foremost objective of this research is to identify the relationship between deforestation rate and the subsequent growth of carbon footprint levels in Indian metropolitan cities. This research also tries to predict the additional vegetation cover required to cope up with the estimated deforestation rate and carbon footprint growth over the next 25-years. The block diagram of the proposed methodology is depicted as in Fig.1.

The dataset is collected from various sources which include Indian Government websites as well as NGO's. The acquired data is preprocessed to remove missing values. A relationship between Vegetation, Carbon emission and Population is computed, which helps us to determine the pattern between vegetation change and carbon emissions. The above relationship is then used to estimate the afforestation efforts required for sustenance.
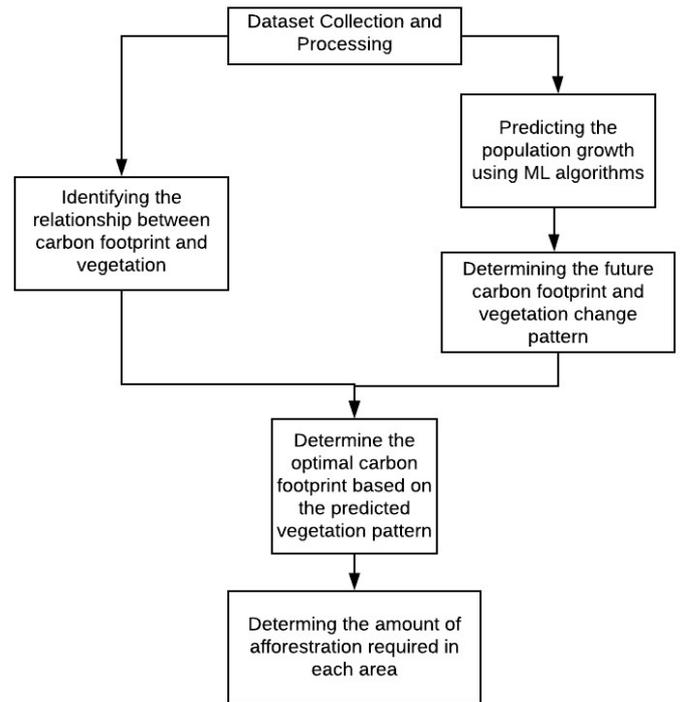

Fig. 1.   Block diagram

IV.   DETAILED ARCHITECTURE

This section provides a detailed explanation of the proposed idea for mapping of carbon footprint with population and vegetation cover in metropolitan regions of India. The detailed architecture of the proposed design is shown in Fig. 2.

A. Dataset Collection

This study involves the collection of data concerning vegetation cover and carbon footprint in Indian cities over the past few decades. The proposed model uses a dataset consisting of land-use statistics of India compiled from various Government sources as well as research studies specified in Table.2. Table.2 also contains details regarding the tree cover loss and carbon footprint datasets. This data encompasses the period from 2000 to 2018. Air quality data with respect to greenhouse gas emissions is collected from the Pollution Control Board's website. Multivariate Imputation by Chained Equations (MICE) is used to fill missing values by taking the multivariate nature of data into consideration. Table II. of the study provides a detailed description of data collection sources.

TABLE II.   DATASET DESCRIPTION

Category	Subcategory	Data source
Land Use Statistics		
District-wise, 28 states and 7	Non-agricultural	Directorate of Economics And

union territories (Area in Hectares)	land, Barren areas, Unculturable Land, Forest Area, Geographical Area, Culturable Waste Land, Land Under Misc. Tree Crops, Mangroves, Net Sown Area, Current Fallows, Permanent Pastures, Grazing Lands	Statistics, Govt. of India(1998 - 2009) [12]
		Department of Agriculture, Govt. of Maharashtra (2010 - 2018) [13]
	Tree cover loss	The Global Forest Watch organization (2000 - 2018) [15, 17, 18, 19]
Carbon Footprint		
District-wise, 28 states and 7 union territories (CO2 in metric tons)	Carbon dioxide emissions equivalent to national biomass loss	The Global Forest Watch organization (2000 - 2018) [15, 17, 18, 19]
According to air quality monitoring stations, 28 states and 7 union territories (concentrations in mg/m3)	Air quality with respect to greenhouse gas concentrations (Benzene, CO, Eth Benzene, MP Xylene, NH3, NO, NO2, SOx, NOx, Ozone, PM10)	Central Pollution Control Board, Ministry of Environment, Forest and Climate Change, Govt. of India [16]

B. Data Preprocessing

The final integrated dataset consists of 6 spatial-temporal features, namely - State, District, Vegetation Cover, % Vegetation Cover, Population and CO2 emissions (in metric tons.) from 1998 to 2018.

The data for forest area, pasture land, orchid plantations and miscellaneous tree crops have been considered in computing vegetation cover. The percentage of vegetation cover (%V) is calculated using the formula:

$$\%V = \frac{Vegetation\ Cover}{Total\ Area} \times 100$$

Where, %V is a useful parameter for districts with fluctuating geographical boundaries. Missing values in the dataset are handled using techniques such as K Nearest Neighbours (KNN) imputation and Multiple Imputation with Chained Equations (MICE). The resulting dataset is a panel data with different districts as the cross-sections.

C. Goals

This study aims to achieve the following goals:

1) Afforestation

Identifying the relationship between carbon emission rate and deforestation can help in determining the additional vegetation cover needed to curb the carbon footprint growth. This study can furthermore suggest sustainable techniques and optimal places for afforestation in metropolitan cities.

2) Biodiversity

The biodiversity in metropolitan regions is rapidly declining due to constant deforestation. The implementation of suggested afforestation measures will help to improve biodiversity considerably. Restoration of ecosystems and habitats will help to protect several endangered species and prevent their extinction.

3) Retaining Groundwater

Trees have a fantastic capacity to assimilate and hold water. It is estimated that, as much as 1,000 tons of precipitation can be absorbed by a mature tree. However, in the absence of trees, a lot of energy is required for pumping and filtration.

4) Carbon Sequestration

As trees utilize carbon dioxide to prepare their food, afforestation can help to control CO2 levels. A research done by EPA indicates that the shade and water vapor discharged by trees help to reduce temperatures in cities by as much as 25°C.

D. Prediction and Forecasting models

This study consists of three distinct phases. The first phase is to identify how carbon footprint growth is associated with deforestation. The second phase estimates the additional vegetation cover required over the next 25 years to cope up with the increase in carbon footprint and deforestation during that period. The third phase of this research suggests optimal techniques for sustainable afforestation according to region-specific needs. The component diagram of the proposed methodology is shown in figure 2.

1) Modelling of Carbon Footprint and Deforestation relationship (CDR)

Each feature in the final dataset can be considered as an independent time series. Techniques such as Autocorrelation function (ACF) and Partial Autocorrelation function (PACF), Causality Tests, etc. can be used to quantify the synchrony between the vegetation change (Vi) and carbon footprint/ CO_2 emission (Ci) time series.

The k-NN (k - Nearest Neighbours) Intersection Algorithm can also be used to identify the relationship between Vi and Ci. The idea of this algorithm is that if an observation has k nearest neighbours in one vector representation and the same nearest neighbours in the other, the two vector representations are likely to be related. This likelihood can be calculated by the overlap of the nearest neighbours collectively for all the observations. DTW is used as the distance measure in k-NN Intersection Algorithm. [20]

Granger Causality test will be used to determine whether the time series of vegetation cover and population are useful in the forecasting of CO2. This will help to understand the cause and effect relationship between the features.

2) Estimation of required vegetation cover (Vr)

CDR identified in the first phase serves as an input in the estimation process. This phase also involves the prediction of carbon footprint growth (Cp) and variation in vegetation cover (Vp) over the next 25 years, which is based on historical time series data and population estimates.

The yearly optimal carbon footprint (Co) can then be determined for each region based on Vp and CDR. The difference between Cp and Co measures the carbon footprint (Ca) that cannot be negated by the estimated vegetation cover (Vp). Some amount of afforestation (Vr) is required to deal with Ca. Vr required for each district can be predicted with CDR and Ca.

Time series forecasting techniques such as Auto-Regressive Integrated Moving Average - X (ARIMAX) and Vector Autoregression (VAR) models will be constructed for forecasting of Cp with Vp and population as exogenous variables. LSTM based neural network and Random Forest algorithm will be used for machine learning-based estimation. Panel data estimation will be performed using Generalized Estimating Equations (GEE).

Results of the above-mentioned techniques will be evaluated and the best performing method will be selected as the forecasting model.

3) AI-based suggestions for sustainable afforestation

The objective of this phase is to provide region-specific recommendations for afforestation, based on the estimated Vr. These recommendations include tree species, forestation techniques and suitable places to carry out afforestation activities.

V. FUTURE SCOPE AND CONCLUSION

This study addresses the optimization and management of tree ecosystems in major metropolitan cities. Considering the growing population of metropolitan cities, the aim is for sustenance along with development. The above-proposed idea will be implemented in the future and develop a system for suggesting suitable areas for afforestation. A system for recommending tree species according to the terrain of the city can be coupled with the proposed system for maintaining the ecosystem.
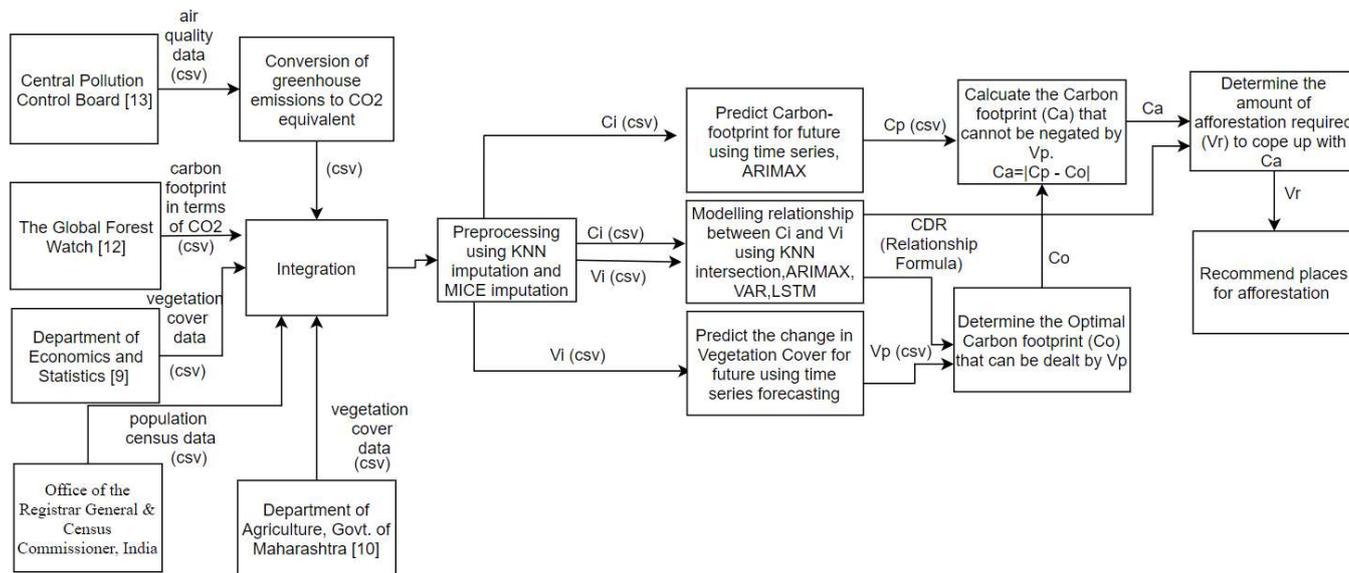

Fig. 2. Conceptual diagram

VI.   REFERENCE

[1] "India's CO_2 emissions are growing faster than the US' or China's" Available from: https://qz.com/india/1581665/indias-carbon-emissions-growing-faster-than-us-china-says-iea/

[2] "Has India's Forest Cover increased over the years?" Available from: https://factly.in/has-indias-forest-cover-increased-over-the-years/

[3] "Bengaluru: One tree for seven people, says an IISc study". Available from: https://www.deccanchronicle.com/lifestyle/pets-and-environment/220416/bengaluru-one-tree-for-seven-people-says-an-iisc-study.html

[4] "India:Environmental Issues" Available from: http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637920/EPRS_BRI(2019)637920_EN.pdf

[5] "India State Of Forest Report 2019 Volume-II":http://164.100.117.97/WriteReadData/userfiles/ISFR2019%20Vol-II.pdf

[6] "India State Of Forest Report 2019 Volume-I":http://164.100.117.97/WriteReadData/userfiles/ISFR2019%20Vol-I.pdf

[7] Peter Walder, Franz Sinabell, Fabian Unterlass, Andreas Niedermayr, Denisa Fulgeanu, Martin Kapfer, Michael Melcher and Jochen Kantelhardt "Exploring the Relationship between Farmers' Innovativeness and Their Values and Aims" DOI: https://doi.org/10.3390/su11205571

[8] Kuldip Singh Sangwan, Vikrant Bhakar, Vinti Arora and Prem Solanki "Measuring carbon footprint of an Indian university using life cycle assessment" DOI: 10.1016/j.procir.2017.11.111

[9] Ben Moews, J.Michael Herrmann and GbengaIbikunle "Lagged correlation-based deep learning for directional trend change prediction in financial time series" DOI: https://doi.org/10.1016/j.eswa.2018.11.027

[10] RakeshKumar, S.Nandy, ReshuAgarwal and S.P.S.Kushwaha "Forest cover dynamics analysis and prediction modeling using logistic regression model" DOI:https://doi.org/10.1016/j.ecolind.2014.05.003

[11] Chairul Saleh, Nur Rachman Dzakiyullah, Jonathan Bayu Nugroho "Carbon dioxide emission prediction using support vector machine" DOI:10.1088/1757-899X/114/1/012148

[12] Directorate of Economics And Statistics, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Govt. of India. District wise land use statistics. Available from: https://eands.dacnet.nic.in/

[13] Department of Agriculture, Govt. of Maharashtra. Land Utilization Statistics. Available from: http://krishi.maharashtra.gov.in/1074/Land-Utilization-Statistics

[14] Hanqin Tian, Kamaljit Banger, Tao Bo, Vinay K. Dadhwal. "History of land use in India during 1880–2010: Large-scale land transformations reconstructed from satellite data and historical archives". DOI: https://doi.org/10.1016/j.gloplacha.2014.07.005

[15] Global Forest Watch. Forest Monitoring. Available from: https://www.globalforestwatch.org/

[16] Central Pollution Control Board, Ministry of Environment, Forest and Climate Change, Govt. of India. Air quality data. Available from: https://www.cpcb.nic.in/

[17] Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." Science 342 (15 November): 850–53. Data available on-line from: http://earthenginepartners.appspot.com/science-2013-global-forest.

[18] Zarin, D., Harris, N.L. 2016. Can carbon emissions drop by 50% in five years? Global Change Biology, 22: 1336-1347. doi:10.1111/gcb.13153

[19] Global Administrative Areas Database, version 3.6. Available at http://gadm.org/

[20] Arighna Roy "Quantifying relationships between two time series data sets". Available from: https://pdfs.semanticscholar.org/8824/fb4ef14f94f96a9d6799551e99a612fa7cf3.pdf