Using Satellite Images and Deep Learning to Measure Health and Living Standards in India

Downloaded from: https://research.chalmers.se, 2026-04-04 23:25 UTC

Citation for the original published paper (version of record):
Daoud, A., Jordán, F., Sharma, M. et al (2023). Using Satellite Images and Deep Learning to Measure Health and Living Standards in India. Social Indicators Research, 167(1-3): 475-505. http://dx.doi.org/10.1007/s11205-023-03112-x

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Social Indicators Research
https://doi.org/10.1007/s11205-023-03112-x

ORIGINAL RESEARCH

Using Satellite Images and Deep Learning to Measure Health and Living Standards in India

Adel Daoud[1,2] · Felipe Jordán[3] · Makkunda Sharma[4,7] · Fredrik Johansson[2] · Devdatt Dubhashi[2] · Sourabh Paul[5] · Subhashis Banerjee[6,7]

Accepted: 3 April 2023
© The Author(s) 2023

Abstract

Using deep learning with satellite images enhances our understanding of human development at a granular spatial and temporal level. Most studies have focused on Africa and on a narrow set of asset-based indicators. This article leverages georeferenced village-level census data from across 40% of the population of India to train deep models that predicts 16 indicators of human well-being from Landsat 7 imagery. Based on the principles of transfer learning, the census-based model is used as a feature extractor to train another model that predicts an even larger set of developmental variables—over 90 variables—included in two rounds of the National Family Health Survey (NFHS). The census-based-feature-extractor model outperforms the current standard in the literature for most of these NFHS variables. Overall, the results show that combining satellite data with Indian Census data unlocks rich information for training deep models that track human development at an unprecedented geographical and temporal resolution.

Keywords Measurement of health and living conditions · Indicators · Deep learning · Satellite images · India · Survey · Census

✉ Adel Daoud
adel.daoud@liu.se

[1] Department of Management and Engineering, Institute for Analytical Sociology, Linköping University, Norrköping, Sweden

[2] The Division of Data Science and Artificial Intelligence of the Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden

[3] Department of Economics and Institute for Sustainable Development, Pontificia Universidad Católica de Chile, Santiago, Chile

[4] Wadhwani AI, Mumbai, India

[5] Humanities & Social Sciences, Indian Institute of Technology Delhi, New Delhi, India

[6] Computer Science, Ashoka University, Sonepat, Haryana, India

[7] Computer Science and Engineering, Indian Institute of Technology Delhi, New Delhi, India

1 Introduction

While country-level data on health and material-living standards—*human development* for short—exist in plentitude for the developing world, within-country data on human development at high spatial and temporal resolution are limited. Such local data, measuring the human development of villages and neighborhoods, are critical for monitoring progress towards the Sustainable Development Goals and enabling tailored public policies to speed up development (Burke et al., 2021; Subramanian et al., 2023).

Recently, scholars have combined earth observations (EO) and machine learning (ML) in developing EO-ML methods to estimate human development using satellite images (Burke et al., 2021; Chi et al., 2022; Head et al., 2017; Jean et al., 2016; Kino et al., 2021; Pandey et al., 2018; Rolf et al., 2021; Suraj et al., 2017; Yeh et al., 2020). Despite the success of EO-ML methods, they are limited in at least four ways: (1) while some studies analyze non-African countries (Chi et al., 2022; Subash et al., 2018; Watmough et al., 2016), most studies focus on African development (Burke et al., 2021; Head et al., 2017; Jean et al., 2016; Yeh et al., 2020), missing how these models work in other regions, such as India, where human development is generally low and unequally distributed; (2) the majority of EO-ML methods focus only on a limited number of outcomes (Head et al., 2017; Yeh et al., 2020), often material assets; (3) EO-ML methods often incorporate nightlight luminosity data as a crude proxy for economic development to boost ML performance (Henderson et al., 2018; Xie et al., 2015), yet other more informative proxies for human development—such as census data—has not been tested yet; and (4) current EO-ML methods rely on cross-sectional data, and it remains unclear how they handle shifts in human-development distribution over time.

This article develops EO-ML methods capable of measuring human development in selected Indian states across time and space at the village level. Specifically, we use Landsat 7 imagery and ML trained on census data. Our methods contribute to addressing the aforementioned knowledge gaps. First, we move beyond Africa and into one of the world's most populous countries: India. Second, although existing studies develop EO-ML methods for only a few outcomes—most notably income and a household asset index—we tailor and evaluate our EO-ML methods for over 90 outcomes that capture many dimensions of human development. Third, instead of relying on nighttime light luminosity (Xie et al., 2015), our method uses transfer learning based on a multidimensional asset index constructed from the Indian Census. Our results show that our asset index outperforms nighttime light luminosity as a target to extract relevant features from daytime imagery. Fourth, we explore the capacity of our model to predict outcomes in periods that were not used to train it. Additionally, we test the performance of a set of transformations on the distribution of outcomes designed to correct for shifts in the distributions of outcomes over time.

2 Background

Traditionally, scholars have used census or household surveys to assess human development (Atkinson, 2016; Daoud et al., 2016; Halleröd et al., 2013). A *census* is a comprehensive measurement of the material living standard of all individuals in a population, yet it is conducted infrequently (usually every ten years) and collects a small number of characteristics of the target population (Randall & Coast, 2015). In contrast, although household

surveys cover a wide range of variables that usually include health outcomes, these surveys have limited reliability for local statistical inference. Increasing the frequency of censuses or surveys would cost governments hundreds of millions of US dollars (Atkinson, 2016). Despite an increase in surveying frequency, scholars would still not have a method for traversing back in time to measure the history of human development.

Geostatistical methods offer an alternative to estimate the distribution of several events, such as population density (Tatem, 2017), air quality (Raheja et al., 2021, 2022), energy consumption (Chithaluru et al., 2022; Samriya et al., 2022; Singh et al., 2022), and human development (Alegana et al., 2015; Steele et al., 2017). Often these methods rely on interpolating and extrapolating geo-temporal data. Although these methods are important for geo-temporal estimation, they are limited by the fact that they require geo-temporal data. That is, they require tabular data for many geographical units, distributed over time and space. For human development, such data are often missing, blocking the effective use of geostatistical methods.

As satellite imagery has existed for several decades (Young et al., 2017), this data source provides a promising low-cost alternative to track human development at a granular spatial and temporal resolution (Jean et al., 2016). Twenty years ago, researchers started to explore this avenue by using nighttime light luminosity as a proxy for human development (Chen & Nordhaus, 2011; Doll et al., 2006; Henderson et al., 2012, 2018; Sutton et al., 2007). Recent advances in the overlap of EO and computer science have made significant progress since then, offering an alternative measurement method that combines daytime light and nighttime light satellite images with ML methods to estimate local characteristics of cities, villages, and neighborhoods (Burke et al., 2021; Chi et al., 2022; Head et al., 2017; Jean et al., 2016; Kino et al., 2021; Pandey et al., 2018; Rolf et al., 2021; Suraj et al., 2017; Yeh et al., 2020). Although an EO-ML method requires ground truth data from census or surveys for training, it tends to outperform the use of nighttime light luminosity to measure human development (Jean et al., 2016). Nonetheless, as previously mentioned, existing EO-ML methods are limited in at least four ways.

First, besides a few exceptions (Chi et al., 2022; Subash et al., 2018; Watmough et al., 2016), most of the success of existing EO-ML methods has been based on measuring human development in Africa (Burke et al., 2021; Head et al., 2017; Jean et al., 2016; Yeh et al., 2020). A critical question is whether the success of EO-ML methods is tied to cultural, demographic, or economic idiosyncrasies of the African population, and thus, how well EO-ML methods generalize to other parts of the world remains to be evaluated.

Second, most of the existing EO-ML methods have been tested on a limited number of cross-sectional outcomes. For example, both Yeh et al.(2020) and Jean et al.(2016) test their method on household income and an asset index that captures a household's general material-living standards. Although Head et al. (2017) makes progress in extending the EO-ML method to other dimensions of human development, their study focuses on a limited number of countries (Nepal, Haiti, Nigeria, and Rwanda) and dimensions (e.g., electricity, mobile-phone ownership, child nutrition). Additionally, although Chi et al. (2022) makes significant progress by producing material-wealth estimates of most low- and middle-income countries, this study is limited to income and an asset index. Thus, much remains to be done to evaluate how well an EO-ML method can measure other aspects of human development from space.

Third, because luminosity has been shown to correlate with economic development (Chen & Nordhaus, 2011; Doll et al., 2006; Henderson et al., 2012, 2018; Sutton et al., 2007), Xie et al (2015) developed a transfer learning method to use luminosity as a proxy training data when poverty data is lacking. Incorporating luminosity has been shown to

boost performance even in data-rich situations (Chi et al., 2022; Yeh et al., 2020). While luminosity data is plentiful—that data exists yearly since the 1990s—more research is needed to evaluate the use of more informative proxies. In the case of human development in India, census data may provide such proxy data.

Fourth, existing EO-ML methods face severe limitations when predicting outcomes for periods not represented in training data, as the distributions of outcome variables shift over time. For example, while ownership of a mobile phone was an indicator of material wealth 20 years ago, today, this product is omnipresent and thus less suitable for measuring material wealth (Daoud, 2018; Gordon & Nandy, 2012; Nandy et al., 2016). Adjusting EO-ML methods for distributional changes remains a challenge for effectively predicting outcomes for periods models that have not been trained on.

If scholars had an EO-ML method that captured human development outside of Africa, they would start making progress toward evaluating the validity and reliability of combining ML and satellite images for measuring human development globally, with high frequency and crisp granularity (Burke et al., 2021; Deaton, 2015). India hosts about 1.4 billion of the world's population, yet it lacks consistent temporal and spatial estimates of human development (Alkire & Seth, 2015; Daoud & Nandy, 2019; Subramanian et al., 2023). With increasing validity and reliability for India, EO-ML methods will become more trustable in measuring the trajectories of human development spatially and temporally. For example, these trajectories will reveal how fast villages or neighborhoods are lifting out of poverty and ill health and how well public policies are working in Indian localities.

As delineated in the Introduction, our article addresses these four knowledge gaps. First, it analyzes how well EO-ML methods work in the Indian context, moving beyond Africa; second, it evaluates how well EO-ML methods estimate human development in multiple dimensions, instead of focusing on a few; third, it leverages transfer learning (described in Sect. 3.2) based on a multidimensional asset index constructed from the Indian Census, instead of nighttime light luminosity; and fourth, it uses a variety of distribution shifts methods to handle change over time.

3 Data and Methods

3.1 Data

Figure 1 provides a map of India, displaying the population shares of the states included in this study in grades of red. We restrict our sample to six states where vector data representing the administrative boundaries of villages in the census data were available: Uttar Pradesh, West Bengal, Bihar, Jharkhand, Punjab, and Haryana. These states have around 218,000 villages, about one-third of all Indian villages, and home to 40% of India's population.

Our analysis relies on five sources of data: (1) household data from the 2001 and 2011 Indian censuses to measure demographic and material-living conditions, (2) the Indian National Family Health Survey (NFHS) data from years 2015–16 (called NFHS-4) and 2019–20 (NFHS-5) to measure health outcomes; (3) village-level-administrative boundaries (polygons); (4) satellite imagery from Landsat 7 for the years 2001, 2011, 2016 and 2019; and (5) nighttime light data provided by the Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) for 2011.

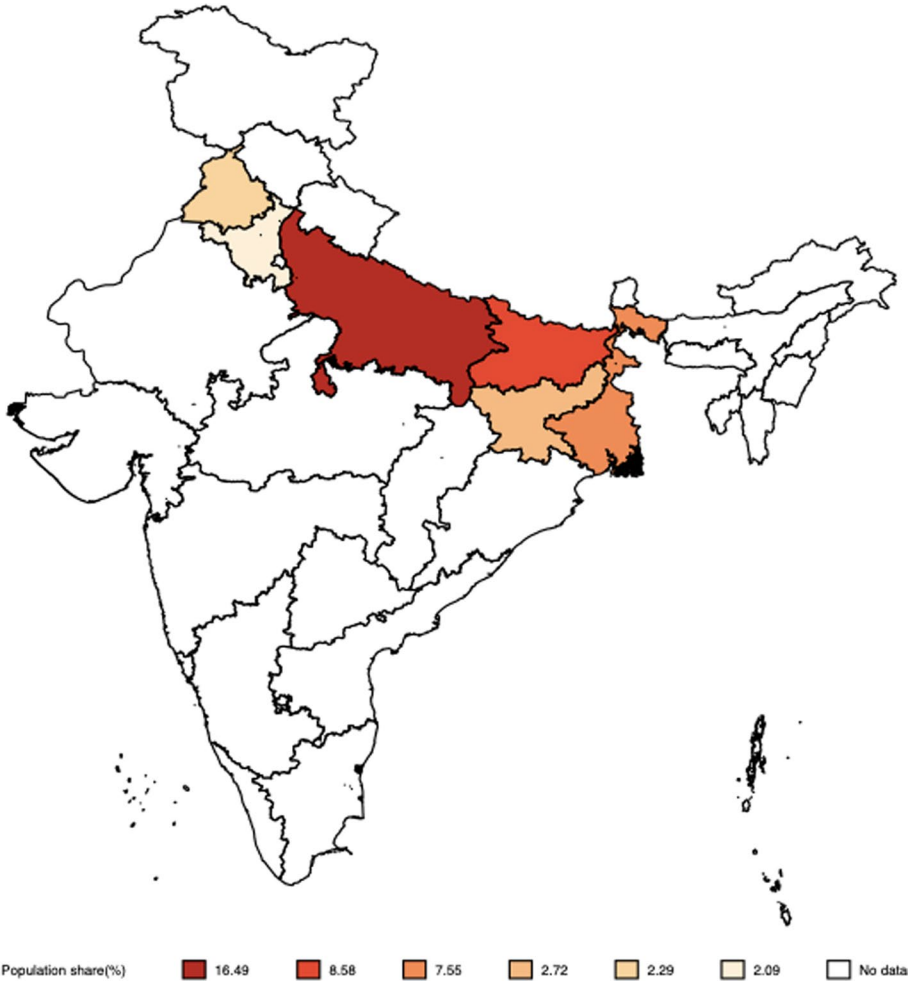


Fig. 1 Scope of the study demographically and geographically

3.1.1 Census and the National Family Health Survey

Our study uses the 2001 and 2011 Indian censuses to collect socioeconomic variables. From the 2011 census, we created 16 asset indices measuring material-living standards, defined in Table 1. For each index, we calculate the average occurrence of material assets in a village. For example, "electronics" is the average occurrence of "radio, transistor, tv, and laptops."[1] The 16-dimensional material-asset vector of outcomes from the census is denoted Y^c.

[1] The underlying material asset is equaly weighted in the aggregation, yet a task for future research is to weight these assets depending on their contribution to human development.

Table 1 The dimension-reduced village-level asset vector derived from the House listing and Housing Census Data, 2011

Variable	Description (percentage of households in the village)	Source (aggregation from columns in the census)[a]
Rooms-under-3	Having less than 3 rooms in the house	[49] + [50] + [51]
Household-size-under-5	Having less than 5 members in the household	[56] + [57] + [58] + [59]
Water-treated	Having access to water from treated source/covered well/ tube-well	[72] + [74] + [77]
Water-untreated	Having access to water from untreated source/uncovered well	[73] + [75]
Water-natural	Having access to drinking water from ponds/rivers/lakes	[76] + [78] + [79] + [80] + [81]
Electric-like	Having access to electricity from grid/solar	[85] + [87]
Oil-like	Having access to fuel from kerosene/other oil	[86] + [88] + [89]
Electronics	Having possession of radio/transistor/tv/laptop	([128] + [129] + [130] + [131])/3
Has-phone	Having possession of land-line/mobile/both	[132] + [133] + [134]
Transport-cycle	Having possession of bicycle	[135]
Transport-motorized	Having possession of motorcycle/scooter/car/jeep	[136] + [137]
No-assets	Having no asset (cycle/phone etc.)	[139]
Banking-services-availability	Availing banking services	[127]
Cook-fuel-processed	Having possession of LPG/electric stove etc	[113] + [114] + [115]
Bathroom-within	Having bathroom within premises	[103] + [104]
Permanent-house	Having permanent house	[140]

[a]The numbers in the square brackets [] denote the column numbers of the Census data table HLPCA-H14, Percentage of Households to Total Households by Amenities and Assets (India & States/UTs)-Village and Ward Level, Houselisting and Housing Census Data–2011

Table 2 The dimension-reduced village-level asset vector derived from the Census Data, 2001

Variable	Description (proportion of households in the Tehsil)	2011 Census		2001 Census	
		Aggregati from columns[a]	Census data table	Aggregation from columns[a]	Census data table
Electric-like	Having energy from grid/solar	([9]+[11])/[8]	HH-7-households by main source of lighting	([3]+[5])/[2]	H-9–distribution of households by source of lighting
Oil-like	Having access to fuel from kerosene/other oil	([10]+[12])/[8]	HH-7-households by main source of lighting	([4]+[6])/[2]	H-9–distribution of households by source of lighting
Electronics	Having radio/transistor/tv/computer	(([10]+[11]+[12]+[13])/3+[20])/[8]	HH-12-number of households availing banking services and number of households having each of the specified sssets	([4]+[5])/[2]	H-13–number of households availing banking services and number of households having each of the specified asset
Has-phone	Having telephone/mobile	([14]+[15]+[16])/[8]	HH-12-number of households availing banking services and number of households having each of the specified assets	[6]/[2]	H-13–number of households availing banking services and number of households having each of the specified asset
Transport-cycle	Having bicycle	[17]/[8]	HH-12-number of households availing banking services and number of households having each of the specified assets	[7]/[2]	H-13–number of households availing banking services and number of households having each of the specified asset
Transport-motorized	Having motorcycle/scooter/car/jeep	([18]+[19])/[8]	HH-12-number of households availing banking services and number of households having each of the specified assets	([8]+[9])/[2]	H-13–number of households availing banking services and number of households having each of the specified asset
No-assets	Having no assets (cycle/phone etc.)	[21]/[8]	HH-12-number of households availing banking services and number of households having each of the specified assets	[10]/[2]	H-13–number of households availing banking services and number of households having each of the specified asset

Table 2 (continued)

Variable	Description (proportion of households in the Tehsil)	2011 Census		2001 Census	
		Aggregati from columns[a]	Census data table	Aggregation from columns[a]	Census data table
Banking-services-availability	Availing banking services	[9]/[8]	HH-12-number of households availing banking services and number of households having each of the specified sssets	[3]/[2]	H-13–number of households availing banking services and number of households having each of the specified asset
Cook-fuel-processed	Having LPG/electric stove etc	([14]+[15])/[9]	HH-10–households by availability of separate kitchen and type of fuel used for cooking	([8]+[9])/[3]	H-10–distribution of households by availability of bathroom type of latrine within the house and type of drainage connectivity for waste water outlet table
Bathroom-within	Having bathroom within premises	[9]/[8]	HH-10–households by availability of separate kitchen and type of fuel used for cooking	[3]/[2]	H-11–distribution of households by availability of separate kitchen and type of fuel used for cooking

[a] The column numbers in the Census data tables are given in square brackets []

As shown in Table 2, while the 2001 census has similar definitions for most variables compared to 2011, some of them are measured at a different level of aggregation. This mismatch in aggregation makes it challenging to compare outcomes across the two census rounds. For example, in 2011, the village-level data report the percentage of households having a telephone. In comparison, the 2001 data uses a binary variable indicating whether a telephone is available in the village. Because of this mismatch, we cannot directly compare these sets of outcomes in the 2001 and 2011 censuses at the village level.

A redeeming aspect of the 2001 Census is that the values for 10 of the 16 variables we construct using 2011-census data are available as a fraction of the population at the tehsil level, one administrative level up from the village level. Table 2 shows how these ten variables are constructed from 2001 and 2011 tehsil-level census data. Although a temporal evaluation for some variables is not possible at the finest level of aggregation (village level), it is still possible if we accept an aggregation at the tehsil level.[2] The methods section delineates our aggregation procedure.

Our study also includes other demographic data from a separate section of the census, denoted by Y^{c*}. It includes these data in the transfer learning models to assess whether the feature extracted when using Y^c as an outcome can be used to predict demographic variables that are likely to display a weaker correlation with satellite imagery's visible features.

While the census measures material-living standards, the NFHS captures mainly health outcomes. The NFHS is a multi-round survey providing information on the health of women and children. Our study uses 93 outcomes denoted by the vector Y^s, where the superscript s stands for *survey*. These outcomes are defined in Table 3. As NFHS surveys do not release information to identify households' villages, we aggregate predictions to the district level for model evaluation, as detailed in the methods section.

3.1.2 A Primer on Image Data

Before discussing our Landsat satellite image source, we discuss the fundamental structure of image data. Image data consists of a squared grid that is composed of pixels. Mathematically, that grid of pixels is a matrix. In a black-and-white image, there is only one matrix where each pixel takes a value representing the intensity of blackness. This matrix represents a band (also known as a channel). A band consists of numeric values, where each value represents radiation (light intensity) within a range of wavelengths of the electromagnetic spectrum (e.g., in Landsat 7, the red band measures radiation between wavelengths of 630 to 790 nanometres). A color image, including a daylight satellite image, consists of pixels populating three bands, thus three matrices, representing the colors red, green, and blue (RGB).

The size of the squared grid—thus the size of the matrix—is defined in height and width that covers a prespecified spatial area. For satellite images, the image size is commonly described in terms of meters or kilometers (km). Our satellite images capture a 3.36-by-3.36 km of each village, and the resolution of that image consists of pixels where each pixel populates a spatial size of 15-m-by-15 m. This means that the vertical side of the square grid is 224 pixels (3360 m divided by 15 m) and the horizontal side consists of

[2] India is divided into the following administrative units, starting from the finest to the most aggregated: a village comprises around 1000 households; a tehsil covers about 100 villages; a district (with an average population of 1.1 million) has around ten tehsils; and a state captures a set of districts. Our data has 189 districts in six states in Northern and Eastern India.

Table 3 The dimension-reduced NFHS district level health vector

Variable	Description
Factor-1	Population (female) age 6 years and above who ever attended school (%)
Factor-2	Population below age 15 years (%)
Factor-3	Sex ratio of the total population (females per 1000 males)
Factor-4	Sex ratio at birth for children born in the last five years (females per 1000 males)
Factor-5	Children under age 5 years whose birth was registered (%)
Factor-6	Households with electricity (%)
Factor-7	Households with an improved drinking-water source1 (%)
Factor-8	Households using improved sanitation facility2 (%)
Factor-9	Households using clean fuel for cooking3 (%)
Factor-10	Households using iodized salt (%)
Factor-11	Households with any usual member covered by a health scheme or health insurance (%)
Factor-12	Women who are literate (%)
Factor-13	Men who are literate (%)
Factor-14	Women with 10 or more years of schooling (%)
Factor-15	Women age 20–24 years married before age 18 years (%)
Factor-16	Men age 25–29 years married before age 21 years (%)
Factor-17	Women age 15–19 years who were already mothers or pregnant at the time of the survey (%)
Factor-18	Any method4 (%)
Factor-19	Any modern method4 (%)
Factor-20	Female sterilization (%)
Factor-21	Male sterilization (%)
Factor-22	IUD/PPIUD (%)
Factor-23	Pill (%)
Factor-24	Condom (%)
Factor-25	Total unmet need (%)
Factor-26	Unmet need for spacing (%)
Factor-27	Health worker ever talked to female non-users about family planning (%)
Factor-28	Current users ever told about side effects of current method6 (%)
Factor-29	Mothers who had antenatal check-up in wthe first trimester (%)
Factor-30	Mothers who had at least 4 antenatal care visits (%)
Factor-31	Mothers whose last birth was protected against neonatal tetanus7 (%)
Factor-32	Mothers who consumed iron folic acid for 100 days or more when they were pregnant (%)
Wfactor-33	Mothers who had full antenatal care8 (%)
Factor-34	Registered pregnancies for which the mother received mother and child protection (MCP) card (%)
Factor-35	Mothers who received postnatal care from a doctor/nurse/LHV/ANM/midwife/other health personnel within 2 days of delivery (%)
Factor-36	Mothers who received financial assistance under Janani Suraksha Yojana (JSY) for births delivered in an institution (%)
Factor-37	Average out of pocket expenditure per delivery in public health facility (Rs.)
Factor-38	Children born at home who were taken to a health facility for check-up within 24 h of birth (%)
Factor-39	Children who received a health check after birth from a doctor/nurse/LHV/ANM/ midwife/ other health personnel within 2 days of birth (%)
Factor-40	Institutional births (%)
Factor-41	Institutional births in public facility (%)

Table 3 (continued)

Variable	Description
Factor-42	Home delivery conducted by skilled health personnel (out of total deliveries) (%)
Factor-43	Births assisted by a doctor/nurse/LHV/ANM/other health personnel (%)
Factor-44	Births delivered by caesarean section (%)
Factor-45	Births in a private health facility delivered by caesarean section (%)
Factor-46	Births in a public health facility delivered by caesarean section (%)
Factor-47	Children age 12–23 months fully immunized (BCG, measles, and 3 doses each of polio and DPT) (%)
Factor-48	Children age 12–23 months who have received BCG (%)
Factor-49	Children age 12–23 months who have received 3 doses of polio vaccine (%)
Factor-50	Children age 12–23 months who have received 3 doses of DPT vaccine (%)
Factor-51	Children age 12–23 months who have received measles vaccine (%)
Factor-52	Children age 12–23 months who have received 3 doses of Hepatitis B vaccine (%)
Factor-53	Children age 9–59 months who received a vitamin A dose in last 6 months (%)
Factor-54	Children age 12–23 months who received most of the vaccinations in public health facility (%)
Factor-55	Children age 12–23 months who received most of the vaccinations in private health facility (%)
Factor-56	Prevalence of diarrhoea (reported) in the last 2 weeks preceding the survey (%)
Factor-57	Children with diarrhoea in the last 2 weeks who received oral rehydration salts (ORS) (%)
Factor-58	Children with diarrhoea in the last 2 weeks who received zinc (%)
Factor-59	Children with diarrhoea in the last 2 weeks taken to a health facility (%)
Factor-60	Prevalence of symptoms of acute respiratory infection (ARI) in the last 2 weeks preceding the survey (%)
Factor-61	Children with fever or symptoms of ARI in the last 2 weeks preceding the survey taken to a health facility (%)
Factor-62	Children under age 3 years breastfed within one hour of birth9 (%)
Factor-63	Children under age 6 months exclusively breastfed10 (%)
Factor-64	Children age 6–8 months receiving solid or semi-solid food and breastmilk10 (%)
Factor-65	Breastfeeding children age 6–23 months receiving an adequate diet10,11 (%)
Factor-66	Non-breastfeeding children age 6–23 months receiving an adequate diet10, 11 (%)
Factor-67	Total children age 6–23 months receiving an adequate diet10,11 (%)
Factor-68	Children under 5 years who are stunted (height-for-age)12 (%)
Factor-69	Children under 5 years who are wasted (weight-for-height)12 (%)
Factor-70	Children under 5 years who are severely wasted (weight-for-height)13 (%)
Factor-71	Children under 5 years who are underweight (weight-for-age)12 (%)
Factor-72	Women whose Body Mass Index (BMI) is below normal (BMI < 18.5 kg/m2)14 (%)
Factor-73	Men whose Body Mass Index (BMI) is below normal (BMI < 18.5 kg/m2) (%)
Factor-74	Women who are overweight or obese (BMI $\sqrt{}$¢,Ä∞¬· 25.0 kg/m2)14 (%)
Factor-75	Men who are overweight or obese (BMI $\sqrt{}$¢,Ä∞¬· 25.0 kg/m2) (%)
Factor-76	Children age 6–59 months who are anaemic (< 11.0 g/dl) (%)
Factor-77	Non-pregnant women age 15–49 years who are anaemic (< 12.0 g/dl) (%)
Factor-78	Pregnant women age 15–49 years who are anaemic (< 11.0 g/dl) (%)
Factor-79	All women age 15–49 years who are anaemic (%)
Factor-80	Men age 15–49 years who are anaemic (< 13.0 g/dl) (%)
Factor-81	Blood sugar level—high (> 140 mg/dl) (%) women
Factor-82	Blood sugar level-very high (> 160 mg/dl) (%) women

Table 3 (continued)

Variable	Description
Factor-83	Blood sugar level-high (> 140 mg/dl) (%) men
Factor-84	Blood sugar level-very high (> 160 mg/dl) (%) men
Factor-85	Slightly above normal (Systolic 140–159 mm of Hg and/or Diastolic 90–99 mm of Hg) (%) women
Factor-86	Moderately high (Systolic 160–179 mm of Hg and/or Diastolic 100–109 mm of Hg) (%) women
Factor-87	Very high (Systolic > = 180 mm of Hg and/or Diastolic > = 110 mm of Hg) (%) women
Factor-88	Slightly above normal (Systolic 140–159 mm of Hg and/or Diastolic 90–99 mm of Hg) (%) men
Factor-89	Moderately high (Systolic 160–179 mm of Hg and/or Diastolic 100–109 mm of Hg) (%) men
factor-90	Very high (Systolic > = 180 mm of Hg and/or Diastolic > = 110 mm of Hg) (%) men
Factor-91	Cervix (%)
Factor-92	Breast (%)
Factor-93	Oral cavity (%)

the same number of pixels. Combining the three bands and the size of the grid together, we obtain an array with three bands, each band consisting of 224-by-224 pixels, and where each pixel in a band takes a numeric value, representing the radiation within the range of electromagnetic wavelengths of that color. By combining the three bands, one obtains an image consisting of an array with dimensions 224-by-224-by-3. Through this combination, objects and patterns emerge in the image. In our case, these patterns refer to roads, vegetation, human settlement areas, and other meaningful entities visible from the sky that correlate with human development.

For each of the 218,000 villages in our sample, we have collected a 224-by-224-by-3 satellite image, denoted as X^d. These raw images constitute the input to the image processing algorithm. Thus, the data consist of pairs of input and output, $\left\{X_i^d, Y_i^c\right\}_{i=1}^n$, where n is the number of villages.

In preparing our pairs of data for EO-ML training, we preprocessed our images. First, the images are divided into batches. Training is more efficient when done in small batches instead of using all the data simultaneously. Second, color pixel values are normalized, thereby, scaling them in the same way. A normalized scale helps the algorithm to find optimal parameters faster. Finally, the image batches are randomly divided into two sets: training and testing. In the training set, the algorithm fits the model. In the test set, the model evaluates its final performance.

3.1.3 Landsat 7 Satellite Images

This article tests the capacity of EO-ML methods to predict human-development-related outcomes from 2001 to 2019, which requires a repository of imagery of consistent quality throughout this period. Because imagery captured by Landsat-7 provides such a repository, our analysis relies on this satellite technology.

Nonetheless, using Landsat data presents several challenges. First, cloud cover limits the availability of imagery, especially in tropical and subtropical regions. Second, black stripes produced by the failure of the Scan Line Corrector found in images collected after May 31, 2003, further limit imagery availability. Third, topographic effects introduce large

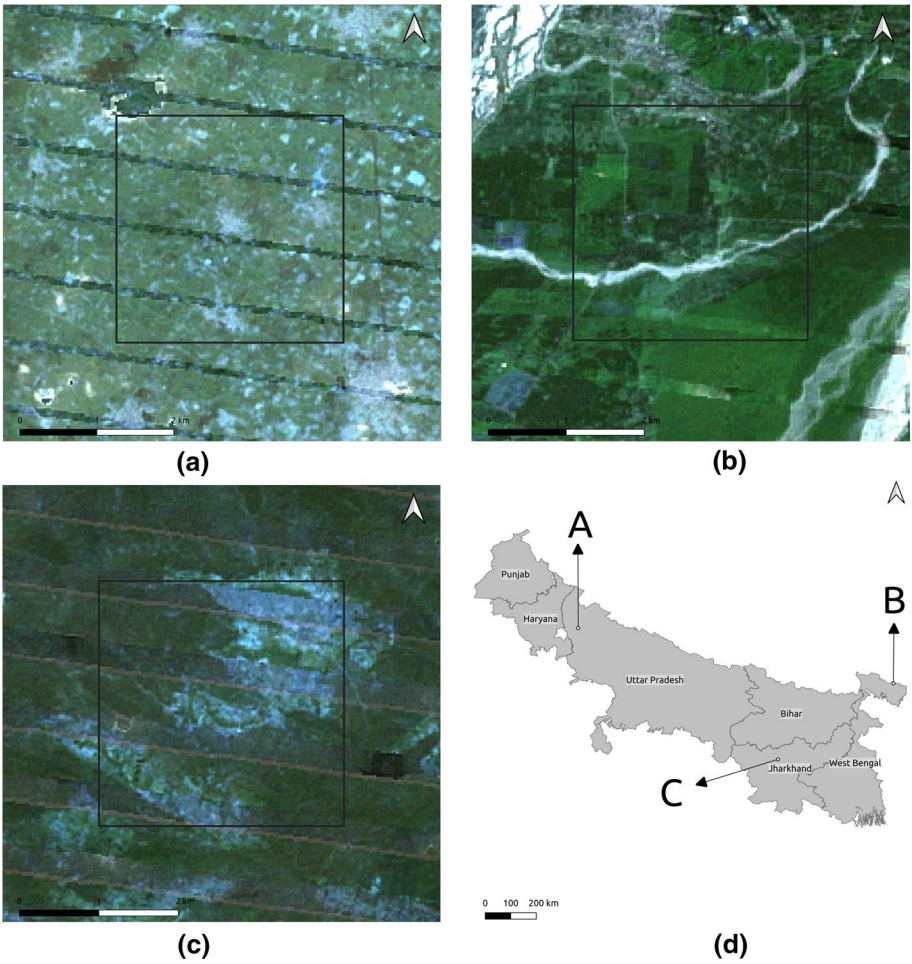


Fig. 2 **a** Landsat 7 pansharpened and terrain corrected Quality Composite for Village A. **b** Landsat 7 pansharpened and terrain corrected Quality Composite for Village B. **c** Landsat 7 pansharpened and terrain corrected Quality Composite for Village C. **d** Location of Villages

variations in the appearance of land covers across space, hampering ML's capacity to detect land cover classes (Khatami et al., 2016). Fourth, the spatial resolution is too coarse in the visible bands (30 m) as compared to state-of-the-art EO-ML methods that use high-resolution imagery.

This paper tackles these challenges by creating annual composites from the repositories of Landsat 7 imagery available in the Google Earth Engine servers. The procedure is run on Google Earth Engine's servers and includes topographic correction (Ekstrand, 1996; Riano et al., 2003; Richter et al., 2009), downscaling of spatial resolution to 15 m using the panchromatic band, and the aggregation of all available imagery into annual composites to minimize the impacts of clouds and the failure of the Scan Line Corrector. Appendix 1 presents details of this procedure. Figure 2 shows the resulting annual image composites for three villages. Panels (a) through (c) show the satellite images of three selected villages, along with a black square delimitating the 224-by-224 pixels used as inputs to the

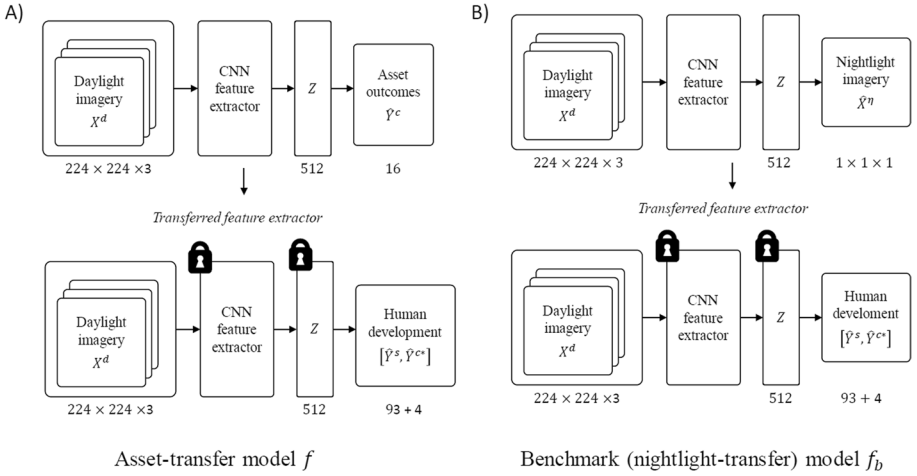


Fig. 3 Deep-learning models. Architecture of two deep models, f and f_b, that utilize transfer learning to predict health outcomes (NFHS) and demographic data (census) from satellite images captured in 2011

deep model. The geographic location of these villages in India is shown in panel (d). While their images may appear to have a medium quality when compared to modern high-resolution imagery, the proposed procedure applies EO's state-of-the-art methods to maximize the quality of annual composites and the results show that they likely contain valuable information to estimate human development.

3.1.4 Nighttime Light Data

While daytime composites capture how villages' material living standards appear from the sky during the day, we use nighttime light data, denoted as X^η, to quantify how much luminosity a village emits during the night. Each pixel contains one band (luminosity value). The more luminosity emitted, the higher the material-living standards tend to be (Chen & Nordhaus, 2011; Doll et al., 2006). Our analysis relies on nighttime light data from DMSP-OLS. The nighttime light data is available in 30 arc-second grids (about 800 m at the latitude of New Delhi), spanning -180 to 180° longitude and -65 to 75° latitude. Each 30-arc-second grid cell is mapped to integer values from 0 to 63, where 63 corresponds to the highest nighttime light intensity.

3.2 Methods

Before providing a primer on EO-ML methods, we provide a birds-eye view of our overarching modeling strategy, as shown in Fig. 3. As displayed in panel (a) of Fig. 3, we will use all the aforementioned data to train a set of EO-ML deep-models f at the village level, using daylight X^d as input for all outcomes. To make efficient use of the few outcome samples in our sample, we use transfer learning. Transfer learning involves solving two tasks: an *auxiliary* task and the *target* task of interest. Here, the auxiliary task amounts to fitting a deep-learning model $f : \boldsymbol{X}^d \rightarrow \boldsymbol{Y}^c$, predicting the census-measured human development

outputs Y^c. The last layers of the trained neural network model f are modified to accommodate the output for the target task, keeping earlier layers (parameters) intact. The task of this modified model is now to predict survey-measured outcomes Y^s and other census demographic outputs Y^{c*}. That is denoted as $f : \boldsymbol{X}^d \rightarrow \boldsymbol{Y}^s \times \boldsymbol{Y}^{c*}$. Here, f is further tuned to predict survey-derived outcomes Y^s and demographic variables Y^{c*}. Our goal with using transfer learning is to let f benefit from the training experience of predicting Y^c and thereby reducing the requirement of collecting a large sample—data points that are expensive to collect or unavailable.

As shown in panel (b) of Fig. 3 and discussed later, we compare our model f with a benchmark model f_b. The benchmark model uses daylight images X^d as input but nightlight data X^n as output in the auxiliary task. Also relying on transfer learning, f_b will then predict survey Y^s and census outcomes Y^{c*}.

3.2.1 A Primer on Machine Learning for Image Data

After pre-processing the satellite images, they are ready to be fed to the image processing algorithms. The input to the algorithm is three matrices representing the three colour bands. There are a variety of image processing algorithms for handling this input (e.g., one of the first algorithm developed is the multilayer perceptron[3]) but those algorithms that perform the best on a variety of image prediction tasks, build on a basic architecture called convolutional neural networks (CNN) (LeCun et al., 1989). By *basic architecture*, we mean algorithmic components (operations) that are shared across modern image processing algorithms. This basic architecture consists of two algorithmic stages: 'identifying a feature representation' and 'conducting the prediction'.

In the feature-representation step, the image processing algorithm estimates which image characteristics (features) are predictive of the outcome—in our case this outcome is the vector of sixteen values that captures the material-assets from the census, Y^c. A feature can be concrete, abstract, and the continuum between concrete and abstract. A concrete feature is for example a visible region of an image such as a road, lake, or human settlement. An abstract feature refers to latent characteristics such as the combination of image regions and patterns not directly apparent to the human eye (Decelle, 2022). In classifying satellite images, the learned latent features often correspond to the nested representation of the image.

The algorithm identifies features by applying three operations sequentially and repeatedly. Because these operations follow each other sequentially, these operations are also called layers. The three operations are the convolution layer, activation layer, and pooling layer. The convolution layer is an operation that quantifies how well all sections of an image match a set of a predefined number of filters—the number of filters depends on the deep-learning architecture (Goodfellow et al., 2016). A filter encodes an image pattern. The intuition is that some filters encode horizontal or diagonal lines, while others measure the prevalence of arches or diagonals. In practice, however, while the number of filters is predefined by the architecture of the image processing algorithm, the encoding of a filter's contents is learned through training the model, and exposing it to a specific dataset. The filters are learned through an estimation procedure called backpropagation (LeCun et al.,

[3] A perceptron takes a column as input. Each band matrix is loosened by stacking all its pixels values into one column. For a 224-by-224 image matrix, the resulting column has 50,176 entries (that is because the matrix has 224*224 = 50,176 entries).

1989, 2015). The overarching task of backpropagation is to populate these filters with numeric weights such that they minimize the sum of squared prediction error; that is the error between the training sample of each village's human development and the model's prediction of human development. The test set is only used to evaluate the performance of the model when all the weights and other parameters have been fully specified. The test set evaluates the generalization capability of the model.

In the convolution layer, when the algorithm applies a filter to an image segment it applies an operation called *convolution*—hence its name a 'convolution layer' or 'convolution neural networks.' In the context of deep learning, a *convolution* is a mathematical operation that evaluates how well an image section matches a filter. Initially, the backpropagation algorithm populates a filter with random weights (or weights used from pretraining on other images from ImageNet), which then also represents a nonsensical feature. The deep-learning algorithm then convolves that filter over the entire image, striding over its region by region. For each stride, it applies a convolution: it calculates the dot-product between the pixel values of the image segment and the filter weights. The dot product provides a metric of similarity between the filter and the image region; the larger the similarity, the larger the dot-product value. For example, the dot product (denote it as a) between a 3-by-3 filter with weights $w_1 \dots w_9$ (populating the filter row by row) and an image region with pixel values $a_1 \dots a_9$ is $a = a_1 w_1 + \dots + a_9 w_9$. If there is no similarity between the filter and the image segment, the dot product will equal zero; if the similarity is high, the dot product will be large. Because the starting weights are randomly initialized, the mean-squared error (MSE) of the initial model will be poor. But as the backpropagation operation updates the weights, the better filters it will find, thereby lowering the MSE.

As said, convolution is a dot product that produces a linear combination of the weights and the pixel values. However, to capture nonlinear combinations and obtain efficient training (with the help of smooth gradients), this dot product passes through a non-linear function, called an activation function. That function is often the Rectified Linear Unit (ReLU) function. The ReLU function is defined as $\max(0, a)$, which means that if the dot product is less than zero the output of ReLU is zero, otherwise it retains a. This resulting output means that the activation function considers only filter-image-segment similarities that are sufficiently large. To prioritize larger similarities, deep learning models include a bias term in each convolution. For each stride i an intercept b_i is added (known as *bias*) to set a higher threshold for when the ReLU is activated, thereby making the model more conservative for when it activates this part of the parameters space. Thus, each stride i produces a dot product containing the following terms, $a_k = a_1 w_1 + \dots + a_9 w_9 + b_k$. The activation layer applies the activation function to all convolved units k between the filter and the image regions in that deep-model depth.

To retain the original size of the image, the algorithm adds padding to the image; that is an additional boarder of pixels surrounding the entire image. The values of those pixels are often set to zero, defaulting to a black color. Padding preserves the size of the original image and it also enables the image edges more possibility to affect the convolution operator and thereby the activation layer. When padding exists, the output of the convolution and activation layers is a new matrix with the same size as the original image. Those values now populate a processed image, with all the corresponding dot-products $a_1^1, \dots, a_k^1$. For our satellite images, they produce the following number of processed pixels: $k = 224 * 224 = 50,176$ pixels.

The pooling layer reduces the dimension of the output of the activation layer. That reduction also shrinks the number of parameters the model must estimate, demand for data, and thus, counters overfitting. The pooling layer consist often of a 2-by-2 kernel that strides

over the output of the activation layer, computing the maximum. Thus, it is called *max pooling*. In each stride, that kernel calculates the maximum of its 2-by-2 window. That maximum value now populates a new, shrunken image. Although the image is shrunken, the algorithm retains the most salient information in each stride.

In a nutshell, the three operations of the convolution, activation, and pooling layers summarize the information of an image. Those summaries are comparable to other statistical operations—such as the mean or variance—but tailored to image data. That summarization of three operations is often applied repeatedly, which refers to the depth of the architecture, often also called hidden layers. Different modeling architectures have different depths and are suitable for different data sizes.

Nonetheless, regardless of the architecture, the main modeling task of the feature-representation step is to learn the filters by estimating their weights and biases. Once the model learns those filters, it has a suitable feature representation of which features in the satellite images are predictive of human development.

The last step of a deep-learning architecture culminates with a prediction step. That step is conducted by taking the final activations and their weights and passing them to a fully connected layer. The fully connected layer is comprised of its final weights, connecting to the output layer. These outputs can be binary or categorical, in which case the model is predicting classification; or they can be continuous, which would be called regression. In the case of binary or categorical output, the final layers pass the fully-connected-layer weights through a SoftMax function. This function converts the output into probabilities of category membership. In our case, the output layer consists of human development indicators.

Transfer learning is a methodology where a deep learning model is trained on one task (e.g., predict ImageNet categories) and then adapted to another (e.g., predict human development). *Transfer* refers to adapting the weights and biases to a new task. It achieves that by replacing the fully connected layer with a fresh fully connected layer, adapted for a new task (Zhuang et al., 2021). Then, it tends to fix the weights and biases of the hidden layer and estimate only the weights and biases of the fully connected layer. Alternatively, one can fine-tune the weights of all layers, if sufficient data exists.

3.2.2 Our Selection of Machine Learning Algorithms

Our analysis relied on a variety of deep-learning architectures, denoted as a set of functions f. To evaluate model dependency in our experiments, we used the following set of functions: ResNet-18 (f_{18}), ResNet-34 (f_{34}), ResNet-50 (f_{50}), VGG-16 (f_{16}). The Visual Geometry Group (VGG) model is one of the early deep-learning architectures, showing a wide range of applicability (Simonyan & Zisserman, 2015). The number 16 denotes the number of convolution layers the model uses. Residual neural network (ResNet) is a deep learning model that uses the fundamental architecture previously discussed but adds a skip connection, allowing information to flow more efficiently between the depth of the model (hidden layers) (He et al., 2016). A skipping connection is a shortcut where a convolution is passed forwards, deeper into the modeling layers. As for the VGG model, the ResNet numbers 18, 34, and 50 refer to the number of convolutions layers the model contains. The higher the number, the deeper the model.

All models are pre-trained on ImageNet (Deng et al., 2009). That means that the filter parameters (weights and biases) of the hidden layers the models use are not randomly initialized. Then we tune the model for our regression tasks to predict Y^c. While the hidden layers use ImageNet parameters, the fully connected layer starts with random weights. Our

training procedure replaced the last fully connected layer of dimensions 512×1000 of each pre-trained network by a randomly initialized layer of dimensions 512×16. The procedure fine-tuned the pre-trained network and freshly train the randomly initialized last layer for transfer learning of the material-asset vectors, using samples from X^d as input and the corresponding Y^c as output. Our estimation procedure relies on the Adam optimizer with a learning rate 10^{-3} and a batch size of 64. To enhance model-training performance, we compute and use the normalized band-wise mean and standard deviation of our dataset instead of the mean and variance of the ImageNet data. Our estimation use a train-test split of 8:2.

The loss function, L, is based on a composition of the 16 human-development outcomes. The fully connected layer has 16 output layers, one layer for each development outcome. To optimize the EO-ML model simultaneously for all of these outcomes, we calculate the 16 mean sum of squares (MSE) for each output layer, and weight them equally. Mathematically, that loss function is the following expression, $L(f(x), y) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{16} \sum_{c=1}^{16} \left(f^c\left(x_i\right) - y_i^c \right) \right)$ Here, the ML model is $f(x)$ and $f^c\left(x_i\right)$ is that model's MSE prediction for the human-development outcome c.

3.2.3 Transfer Learning from Material-Asset Vector to Demographic and NFHS Data

Our research design relies on two assumptions. The first assumption is that satellite images record human development from the sky. That is, X^d contains predictive information about the material-asset vector Y^c. This assumption is reasonable as household asset indicators have been found to correlate with the observable features in the daytime satellite images like the proportion of built-up area, road area and road types, density and type of housing, water bodies, forest cover, and green areas (Jean et al., 2016).

Our second assumption is that f can be leveraged to indirectly estimate Y^s and Y^{c*} from X^d using transfer learning. While Y^s and Y^{c*} contain outcomes that are likely to display a weaker correlation with imagery's raw features, f is expected to extract abstract features that are better positioned to deliver accurate predictions for these outcomes. Thus, with transfer learning, we can expand the dimensions of human development that can be studied (Daoud et al., 2019; Kino et al., 2021), even in small datasets that do not offer a large-enough training set (Zhuang et al., 2021).

While transfer learning to demographic census data (Y^{c*}) requires no additional steps, transfer learning to NFHS data needs to adjust for the different level of aggregation at which census and NFHS data are available (village vs. district). To address this difference in aggregation, we pass the village Landsat image to our model and extract the layer just before the final prediction layer of the 16-dimensional material-asset vector. Then, we average this layer across all villages in each district, weighing by villages' population. Using this averaged layer as inputs, we train a neural network with two fully connected layers with rectified linear activation to do a regression on Y^s.

3.2.4 Temporal Evaluation

To evaluate the performance of f across time, we perform experiments where we first train our model using 2011 inputs and outputs: $f\left(X^d_{2011,i}\right) = \hat{Y}^c_{2011,i}$. Then, for evaluation we use

held-out 2001-census data, $Y^c_{2001,i}$ and $X^d_{2001,i}$. That is, $f\left(X^d_{2001,i}\right) = \hat{Y}^c_{2001,i}$. Our analysis conducts the temporal evaluation at the tehsil level (indexed by h), where definitions of census variables match for 10 out of the 16 components of the material-asset vector. To conduct this evaluation, we aggregate predictions at the tehsil level using a weighted average of village-level predictions, $\hat{Y}^c_{2011,h} = \sum_i w_i \hat{Y}^c_{2011,i}$, where w_i is the share of teshil's h population living in village i. Thus, the target loss we aim to minimize is the sum of squares over all tehsils, that is, $\sum_h \left(\hat{Y}^c_{2011,h} - Y^c_{2001,h}\right)^2$.

A challenge is that the distribution of some outcome variables across villages experienced a significant shift over time (e.g., ownership of cell phones), whereas the distribution of features of imagery across villages (e.g., roads, constructed area) contain only a modest shift in the same period. This mismatch between input (e.g., satellite images) and output (e.g., cell phones) challenges an approach that aims to directly evaluate ML models across time, and thus, we resort to indirect methods that take into account the temporal shift in the distribution of the outcome variables.

Our analysis tested three distribution transformation, denoted by $g()$, to align 2001 to 2011 ground truth distributions: (i) *Simple transform*, which matches the mean and variance of the 2001 ground truth census data to mean and variance of 2011 census data before evaluation; (ii) *Histogram matching*, which transforms 2001 census data by matching histograms of the 2001 census and 2011 census at 10 bins for each variable*;* and *(iii) Linear optimal transport*, which learns a linear optimal transport from 2001 to 2011 census on the training data, and applied it to 2001 census test ground truth before evaluating (Papadakis, 2015).

3.2.5 Benchmark Model

Our assessment evaluates the relative performance of f, which uses a 16-dimensional material-asset vector derived from census data as a feature extractor, by comparing its performance to one of the current standards in the EO-ML literature, which uses nighttime light data instead (Henderson et al., 2012; Xie et al., 2015). As Fig. 3 shows, this baseline consists of first training an ML model to predict nighttime light values from daytime satellite imagery. The relevant nighttime light cells consist of the latitude and longitude for the centroid of the daytime satellite image.

Relying on transfer learning, we then replace the last prediction layer of the nighttime light model to predict Y^{c*} and Y^s, and compare the R^2 to the ones obtained by the models described in the previous section.

3.2.6 Aggregations

For all levels of aggregation and experiments, the satellite-input data is always collected at the village level X^d_{it}. Yet our analysis relies on different aggregation levels, because of either mismatch in definition at the village level or missing latitude–longitude information for outcomes Y_i. First, all census-based cross-sectional results rely on village-level data i, and thus, do not use any aggregation. Second, for all NFHS experiments, we conduct the evaluation at the district level: although ground-truth data Y^s_i is collected at the village-level, we aggregate to the district level d to calculate the loss function because villages are

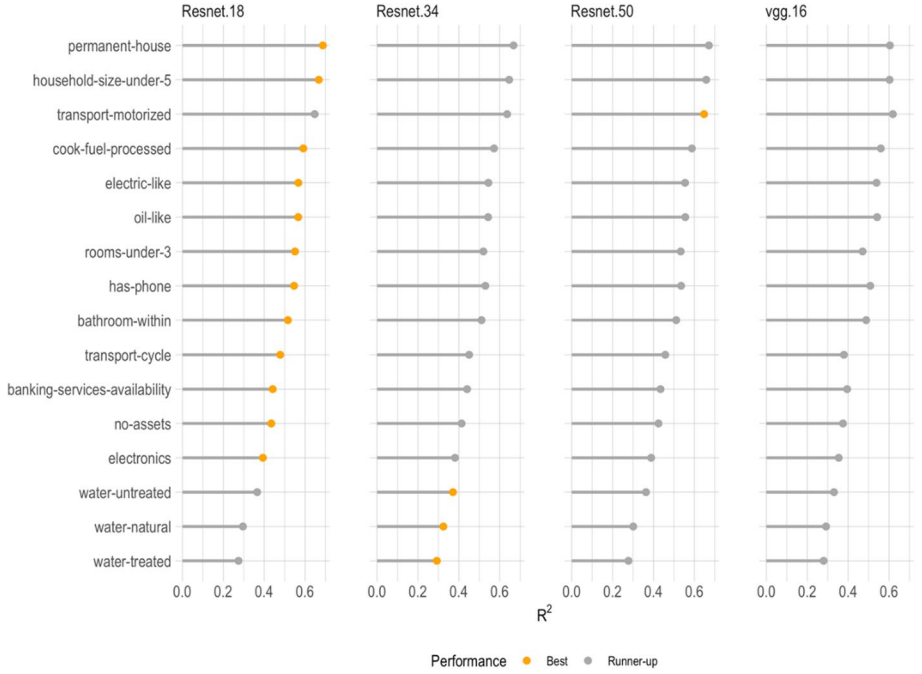


Fig. 4 Relative performance of four deep-learning models, ResNet-18, ResNet-34, ResNet-50, and VGG-16, in prediction of Multidimensional Asset Index

not identified. Third, for the temporal analysis of census data, we conduct the model evaluation at the tehsil level where the definitions of census variables match across time.

4 Results

4.1 Estimating the 16-Dimensional Material-Asset Vector for Transfer Learning

As previously discussed, our analysis relies on four ML models: ResNet-18, ResNet-34, ResNet-50, and VGG-16. As shown in Fig. 4, although all models had comparable model performance, ResNet-18 performed slightly better for 13 of the 16 outcomes. ResNet-34 trails the ResNet-18 performance, and ResNet-34 performs better on water-related outcomes, with R^2 in the range of 0.3 and 0.4. Henceforth, we only present results from ResNet-34, as this is the best overall-performing model: it performs similarly to ResNet-18 on 13 of 16 outcomes, but better than ResNet-18 on the last three outcomes.[4]

Our ResNet-34 (henceforth referred to as f_{34}) model's average R^2-performance, across all 16 outcomes using census 2011, is 0.5, with a standard deviation of 0.12. As the leftmost panel in Fig. 4 shows, our f_{34} performance ranges from as high as $R^2 = 0.69$ (permanent house) to a low of $R^2 = 0.27$ (water-treated). All our models perform well on outcomes that have a physical appearance from the sky (e.g., housing quality), and it tends to

[4] Other model results are available upon request.

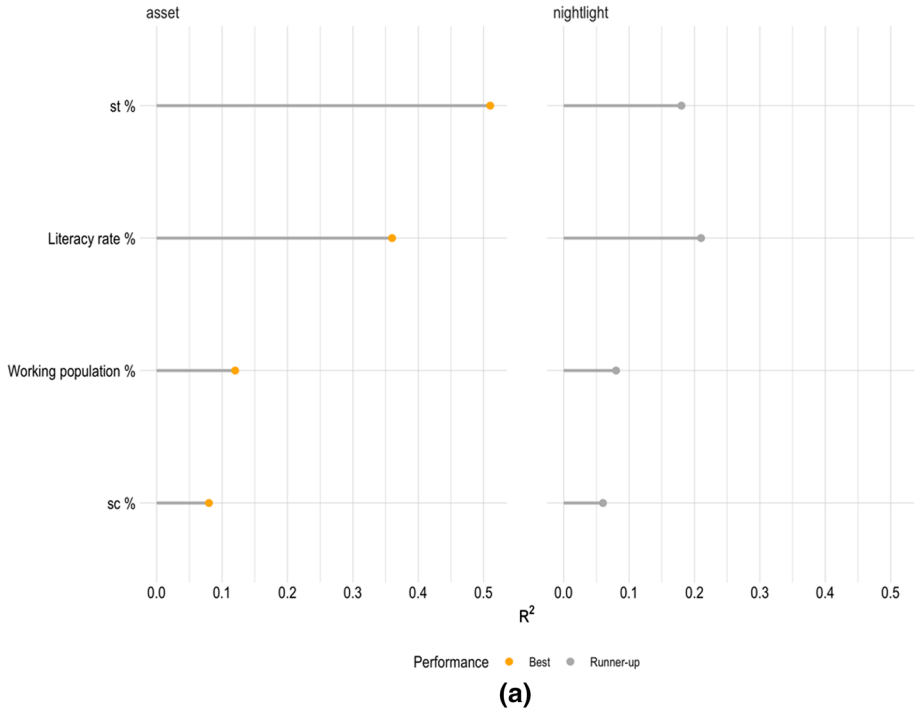


Fig. 5 2011 Multidimensional Asset Index results. **a** Comparison of transfer learning of summary population and demographic data at the village level using the nightlight and asset models. The Population and Demographic variables are extracted from the Population Census Abstract. The results are based on the 16-dimensional asset model to remotely measuring demographic characteristics with transfer learning. st and sc stand for scheduled tribe and scheduled caste, respectively. **b** Comparison of transfer learning of the NFHS-4 variables at the district level using the asset models and 2015 images. The figure captures the R-squared of the 93 NFHS variables. Acronyms: JSY = Janani Suraksha Yojana, ARI = "acute respiratory infection"MCP = Mother and Child Protection Health personnel = doctor/nurse/LHV/ANM/midwife/and similara health personnel

struggle with outcomes that merely correlate with outcomes appearing from the sky (e.g., a household's water-quality access correlates with housing quality).

For the results in Fig. 4, our models are directly learning to associate the daylight-image input, X_i^d to the census-material-living-standards outputs Y_i^c. That is, no aggregation or transfer learning is used. In what follows, we assess if this model can be successfully used to expand the scope of predicted variables using transfer learning.

4.2 Cross-Sectional Transfer Learning Results

In Fig. 5, our model f_{34} relies on transfer learning to predict variables from the demographic section of the 2011 census (Y^{c*}, panel a) or NFHS-4 (Y^s., panels b to d). As discused in the Methods section, the key innovation here is that and Y^s contains outcomes that are distal from what the combination of satellite-images and f_{34} can be expected to predict. For example, as previously mentioned, predicting housing quality from satellite images is

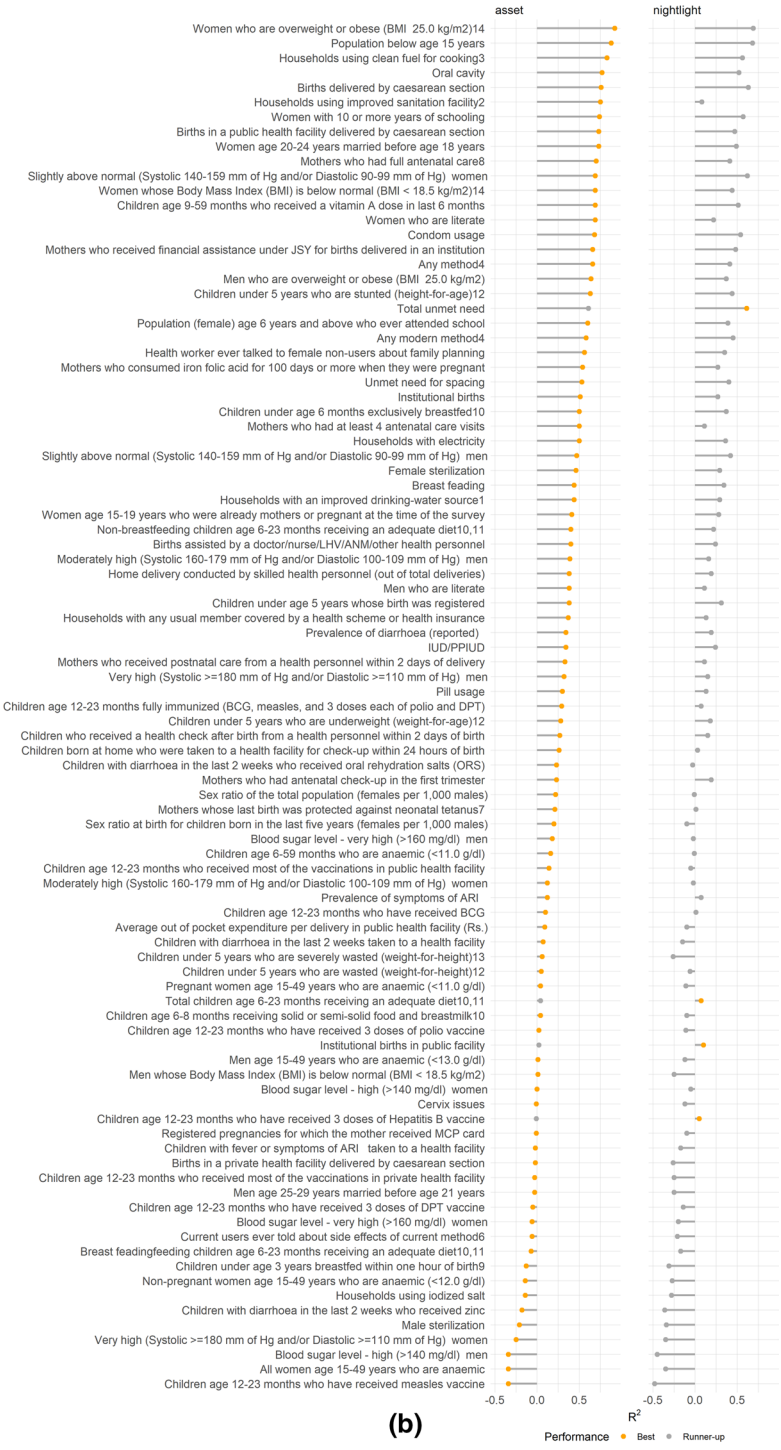


Fig. 5 (continued)

likely to work because roofs, roads, and yards are directly observable from the sky. In contrast, predicting literacy rate or religious affiliation is more challenging, as these outcomes are not directly observable from satellite images.

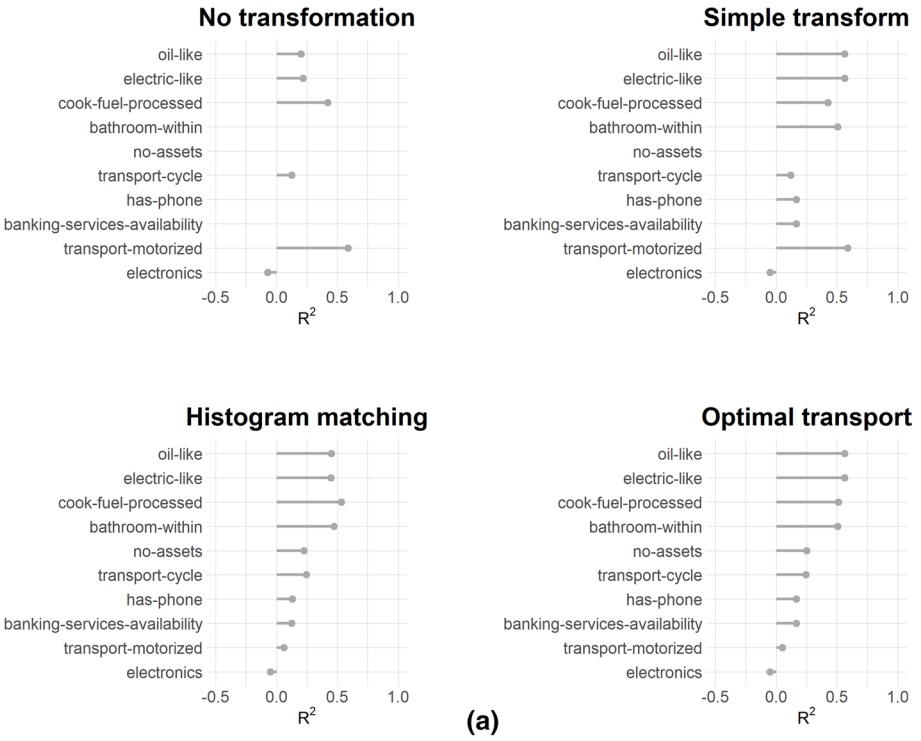
The left-hand side of panel (a) in Fig. 5 shows that our model f_{34}^{*} can predict not only material-living standards but also demographic characteristics with reasonable performance. The model performs best on measuring the share of *scheduled tribe* ($R^2 = 0.49$), followed by *literacy rate* ($R^2 = 0.34$), *working population* ($R^2 = 0.15$), and *scheduled caste* ($R^2 = 0.1$). The right-hand side of the panel (a) in Fig. 5 uses nightlight luminosity in the transfer-learning step instead of our f_{34}^{*} that uses the 16-dimensional material-asset vector for transfer learning. Our f_{34}^{*} performs the best. A nightlight-transfer model has the following performance: *literacy rate* ($R^2 = 0.21$), *scheduled tribe* ($R^2 = 0.18$), working population ($R^2 = 0.08$), and scheduled cast ($R^2 = 0.06$).

Similarly, although Y^s contains mainly health outcomes, f_{34}^{*} has the capacity to predict several of them reasonably well. Panel (b) of Fig. 5 shows the results for transfer-learning based on NFHS-4 outcomes Y^s. The left-hand side of panel (b) shows that our model f_{34}^{*} identifies sufficient signal to predict a variety of health-related outcomes. Of the 93 outcomes, 27 had a score of $R^2 \geq 0.5$. The right-hand side of panel (b) show the results for the benchmark model, which uses nightlight luminosity in the transfer-learning step. Our model outperforms the benchmark in 89 out of the 93 variables. The top five scoring variables when predicting Y^s are *precent of women overweight*, *households with clean fuel*, *share of population below age 15*, *access to condoms for birth control*, and *birth with caesarean section*. The bottom five are *vaccination against measles*, *men with high blood sugar*, *diarrhea treated with zinc*, *women with anemia*, and *women with high blood pressure* (BP). These variables have all negative R^2, which means that the predictions of f_{34}^{*} are worse than just using the sample mean for each Y^s.

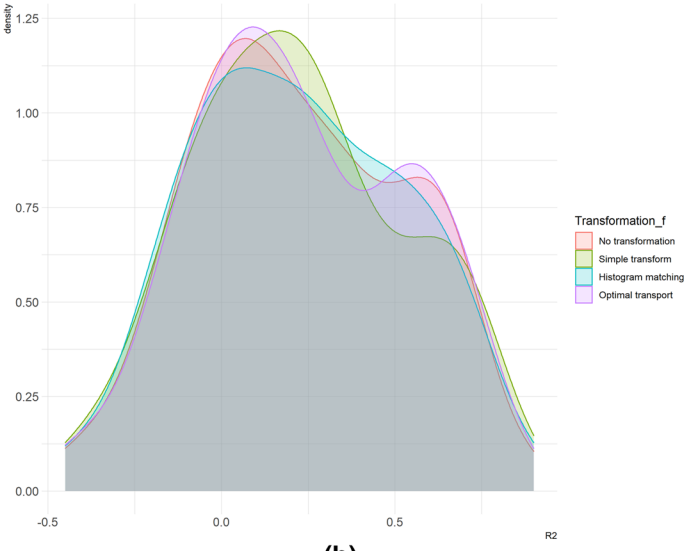
4.3 Temporal Transfer Learning Results

Our f_{34}^{*} trained on the 16-dimensional material-asset vector is able to predict temporal changes, targeting the census 2001 and NFHS-5 (collected in 2019–20). Panel (a), Fig. 6, shows the census-2001 results. As described in the Methods section, besides the non-transformed (original) data, we use three procedures to align the outcome distributions. Using no transformation, the model produces poor results. The worst performing outcome is has-phone, the proportions of phones in villages, with an $R^2 = -236$. The model produces negative R^2 for *electronics* ($R^2 = -0.1$), *banking-services availability* ($R^2 = -2.7$), and *no-assets* ($R^2 = -0.5$). As previously mentioned, negative R^2 means that the model is performing worse than a prediction equal to the sample mean. This is due to shifts in the distribution of outcome variables that outpace changes in the satellite imagery. Next, we show how the three proposed transformations help ameliorate this problem.

While the simple-transform algorithm produces uneven results with negative and positive R^2, the best performing transformations are histogram matching and linear-optimal matching. Based on these two transformations, our model f_{34}^{*} estimates the following outcomes with $R^2 \geq 0.5$: *oil-like* (having energy source from kerosene/other oil), *electric-like* (having energy source from grid/solar), *bathroom-within* (having bathroom within premises), and *cook-fuel-processed* (having LPG/electric stove). Outcomes that f_{34}^{*} tends to estimate less precisely are *electronics* (having possession of radio/transistor/tv/laptop) and *has-phone* (having possession of land-line/mobile/both). One reason why histogram matching and linear-optimal transform perform better than simple transform, is that simple transform


No transformation
Simple transform
Histogram matching
Optimal transport
oil-like
electric-like
cook-fuel-processed
bathroom-within
no-assets
transport-cycle
has-phone
banking-services-availability
transport-motorized
electronics
-0.5
0.0
0.5
1.0
R2

(a)


density
1.25
1.00
0.75
0.50
0.25
0.00
-0.5
0.0
0.5
R2
Transformation_f
No transformation
Simple transform
Histogram matching
Optimal transport

(b)

◀ **Fig. 6** Temporal results. **a** Remotely measuring 2001-census outcomes from 2011-census training and with distribution transformations. The figure shows prediction of the 2001 tehsil level asset vector using 2011 asset model and 2001 images for different temporal transformations of the tehsil level asset distributions. The outcomes "no-assets" and "has-phone" have negative $R2$ exceeding -1 and have thus been dropped in the model "No transformation". **b** Estimating NFHS-5 outcomes. The figure shows the distribution of how well each transformation performs in predicting an outcome in NFHS-5. The x-axis is $R2$ performance and the y-axis is the density of the number of outcomes. The four densities represent untransformed prediction, simple transform, histogram matching, and linear optimal transport. **c** Double-transfer learning (from Census 2011 to NFHS-4, and then to outcomes in NFHS-5) with the selected 3 child outcomes

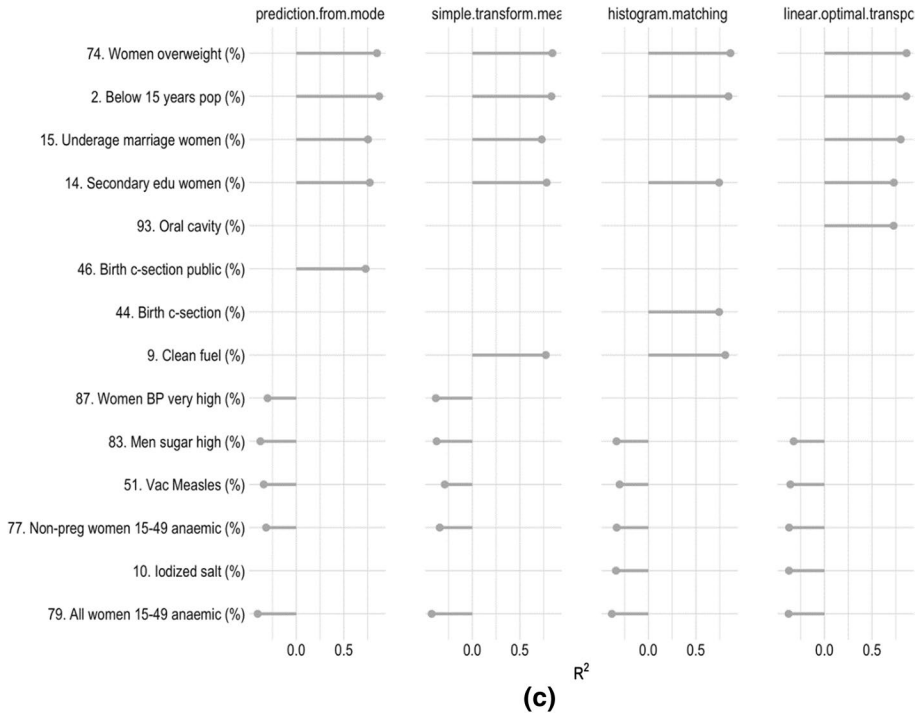


Fig. 6 (continued)

only aligns the two first central moments (mean and variance), while the other two align the 2001 and 2011 census distributions across different distribution characteristics.

Panel (b), Fig. 6, shows the performance of f_{34}^{*} to predicting NFHS-5 (the year 2019) health outcomes (93 variables). Here, we rely on double-transfer learning. The first transfer step consists of f_{34}^{*} predicting NFHS-4 outcomes. The second transfer step consists now of modifying f_{34}^{*} into f_{34}^{**}—that is, we change the last layer in f_{34}^{*} enabling it to predict NFHS-5 outcomes. On top of the double-transfer, we also apply the three proposed transformations to check how much they assist in improving the temporal predictions.

It turns out that there is no added value of conducting an additional transformation when predicting NFHS-5. As the four densities largely overlap, all four transformations are performing equally. This contrasts with the improvements in accuracy delivered by the same transformations when predicting 2001 census outcomes using models trained

in 2011. The difference may be due to the shorter timespan between NFHS surveys (5 years) and NFHS variables changing more slowly through time.

Focusing on "No transformation," the double-transfer model f_{34}^{**} predicts 4 outcomes with R^2-performance above 0.7 (e.g., *Oral cavity*, *Women with 10 or more years of schooling*, and *Women's BMI*), 22 outcomes with R^2-performance above 0.5, 70 outcomes above R^2-performance above 0.1, and 23 outcomes below R^2-performance of 0.

In panel (c), Fig. 6, we focus on NFHS-5 outcomes related to child health only. Regardless of transformations, f_{34}^{**} performs well in predicting mother's access to antenatal care, mothers who consumed iron folic acid for 100 days or more when they were pregnant (%), and share of children underweight.

5 Discussion

The global community has committed to ambitious targets articulated in the Sustainable Development Goals. Although many governments are vigilant in implementing public policies to improve human development for their populations, policymakers lack reliable methods to monitor the effects of their policies at a sufficiently granular level over time and space (Burke et al., 2021). To tackle this lack, scholars are creating innovative methods that capitalize on the predictive accuracy of ML and the visual granularity supplied by EO (Burke et al., 2021; Daoud & Dubhashi, 2023; Jerzak et al., 2023; Rolf et al., 2021). As most of these EO-ML methods focus on Africa (Chi et al., 2022; Suraj et al., 2017), our article creates a comparable method for India.

That is this article's first contribution: while existing EO-ML methods for Africa cover human development for roughly one-seventh of the world population, our method makes progress towards covering an additional one-seventh.

The second and third contributions are that we evaluate how well EO-ML method applies to a multitude of human development indicators and using transfer learning for improved estimation. Our results show that using a census provides a better leverage than nightlight luminosity for transfer learning for these multitude of outcomes. While nightlight luminosity is a frequently used complementary data source to estimate human development (Henderson et al., 2012; Xie et al., 2015), our experiments show that a 16-dimensional asset index performs better. Using this asset index as a leverage, our EO-ML method transferred, with noteworthy accuracy, to estimate a myriad of health outcomes as measured by NFHS.

A fourth contribution of this paper is that our EO-ML method uses outcome-distribution transformations to better estimate temporal change. Because some dimensions of human development change faster—an outcomes such as *access to mobile phones*—than the material shape of neighborhoods as exhibited in satellite images, EO-ML method can struggle in estimating temporal change (Young et al., 2017). When performing transfer learning between Indian censuses (2011 to 2001), our analysis shows that accessible transformations such as histogram matching or linear-optimal transform boost accuracy by several factors. While these transformations helped the prediction in the census, there is a lack of improved performance between NFHS surveys (years 2015–16 to 2019–20). This lack of difference is perhaps due to the temporal lag between NFHS-4 and NFHS-5 being only five years, and health outcomes changing more slowly than assets. Nonetheless, more research is needed to evaluate when transformations help calibrate EO-ML models.

Of the 93 health outcomes and for all transformations our experiments evaluated, about 70 had positive R^2 values, and the best transformation (i.e., linear optimal transport) produced values $R^2 > 0.5$ for 23 outcomes. Conversely, as the temporal results show, night-light-based transfer produced consistently less competitive R^2 values. Thus, our EO-ML method enables scholars and policymakers to measure health outcomes that are not directly observable from the sky. For example, our top-performing outcomes—with R^2 around 0.6—are *precent of women overweight*, *share of population below age 15*, and *birth with caesarean section*, *woman-underaged marriage* and *women with secondary education*.

Improving maternal and child health outcomes are integral part of the Sustainable Development Goals. Although the Indian economy is growing and Indian governments are improving their public-health and anti-poverty policies, much remains for pulling millions out of poverty in the next several decades (Alkire & Seth, 2015; Drèze & Sen, 2013; Reddy & Daoud, 2020; Thorat et al., 2017). For example, India has halved its population-poverty rate from 45.3 (head-count ratio) in 1993 to 21.9% in 2012, yet about 54 million people still live in extreme poverty and with ill-health. To calibrate public policies, policymakers require geo-temporal data for efficient policy targeting (McBride & Nichols, 2016), from modeling urban air quality (Raheja et al., 2021, 2022) to energy flows (Chithaluru et al., 2022; Samriya et al., 2022; Singh et al., 2022) to human development. Our EO-ML method is one critical piece for enabling such efficient targeting (Aiken et al., 2022).

Appendix 1: Landsat 7 Processing

We build a single image for each village i and year t, using Google Earth Engine's Tier 1 and Tier 2 repositories of Landsat 7 daytime imagery (Gorelick et al., 2017).[5] We process these images on Google Earth Engine's servers into batches. In a given batch and year, we use the Quality Assessment Band to remove imagery with more than 5% of cloud coverage over the batch area, defined as the union of all 3.36×3.36 km squares centered at the centroid of the batch's villages' administrative boundaries. Then, we select four bands for processing: the red, green, and blue soil reflectance bands, plus the panchromatic top-of-the-atmosphere band.

The selected images are processed with the C-correction Teillet method to smooth the effect of topography in the imagery (Ekstrand, 1996; Riano et al., 2003; Richter et al., 2009), which has been shown to improve the capacity of EO-ML methods to distinguish among land cover classes (Khatami et al., 2016). The selected imagery is then ordered from the one with the highest level of Normalized Difference Vegetation Index (NDVI) within the batch area to the one with the lowest. Then, a mosaic is built across the batch area through a recursive method that selects, from the first imagery, all the valid pixels and then moves to the second to fill pixels that were covered by clouds or saturated in the first image. This method continues down the list until all pixels are filled, or the end of the list is reached.

The use of year-round composites allows us to maximize the probability of obtaining data for all pixels in subtropical regions with cloudy, wet seasons, and the use of NDVI as a quality measurement maximizes the chances of observing standing crops in agricultural

[5] Tier 1 contains imagery of the highest quality. Tier 2 imagery have the same radiometric standard as that of the Tier 1 imagery, but do not meet Tier 1's geometric specifications.

lands, which is likely to help the model differentiate agricultural land from bare soil. This composite also enables us to fill the gaps produced by the failure of the Scan Line Corrector found in images collected after May 31, 2003.

Finally, we use a simple Multiresolution Analysis (MRA) pan-sharpening method to combine the RGB 30-m resolution bands with the 15-m resolution panchromatic band and create our final 15-m resolution RGB imagery (Vivone et al., 2015). Panels (a) through (c) in Fig. 2 display the final composites for the year 2011 in three selected villages, whose locations are shown in panel (d) of the same figure. Features like fields, roads, and constructed areas are clearly discernible in the composites. Note that, except for the lines with lower NDVI introduced by the failure of the Scan Line Detector, the composites preserve the spatial–temporal context of the imagery. That is, two neighboring pixels have a high chance of being drawn from the same Landsat image, and bands in any given pixels are always drawn from the same image. This is a result of the method used to construct the composites, which contrasts with the simpler approach of using the median value of each band for each pixel, which introduces unnecessary noise to the input imagery.

Before feeding the imagery to the deep learning models presented in the methods section, our procedure normalizes each band to have a mean of zero and a standard deviation of one across all image samples X^d. When unpacking the batch, the result contain a sequence of $224 \times 224 \times 3$ input tensors, each of which is associated with a single village-year, covering 11.3 square kilometers. This geographical-image size is suitable as it is directly compatible with established deep-learning-model architectures that our analysis relies on (Krizhevsky et al., 2017).

Funding Open access funding provided by Linköping University.

Data Availability Replication code is available at Github, https://github.com/AIandGlobalDevelopmentLab/EOML-for-India.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aiken, E., Bellue, S., Karlan, D., Udry, C., Blumenstock, J. E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. *Nature*. https://doi.org/10.1038/s41586-022-04484-9

Alegana, V. A., Atkinson, P. M., Pezzulo, C., Sorichetta, A., Weiss, D., Bird, T., Erbach-Schoenberg, E., & Tatem, A. J. (2015). Fine resolution mapping of population age-structures for health and development applications. *Journal of the Royal Society Interface, 12*, 20150073. https://doi.org/10.1098/rsif.2015.0073

Alkire, S., & Seth, S. (2015). Multidimensional poverty reduction in India between 1999 and 2006: Where and how? *World Development, 72*, 93–108. https://doi.org/10.1016/j.worlddev.2015.02.009

Atkinson, T. (2016). Monitoring global poverty: Report of the commission on global poverty. *The World Bank*. https://doi.org/10.1596/978-1-4648-0961-3

Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. *Science, 371*, abe8628. https://doi.org/10.1126/science.abe8628

Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences, 108*, 8589–8594. https://doi.org/10.1073/pnas.1017031108

Chi, G., Fang, H., Chatterjee, S., & Blumenstock, J. E. (2022). Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences USA, 119*, e2113658119. https://doi.org/10.1073/pnas.2113658119

Chithaluru, P., Al-Turjman, F., Kumar, M., & Stephan, T. (2022). MTCEE-LLN: Multilayer threshold cluster-based energy-efficient low-power and lossy networks for industrial internet of things. *IEEE Internet of Things Journal, 9*, 4940–4948. https://doi.org/10.1109/JIOT.2021.3107538

Daoud, A., Halleröd, B., & Guha-Sapir, D. (2016). What is the association between absolute child poverty, poor governance, and natural disasters? A global comparison of some of the realities of climate change. *PLoS ONE, 11*, e0153296. https://doi.org/10.1371/journal.pone.0153296

Daoud, A., Kim, R., & Subramanian, S. V. (2019). Predicting women's height from their socioeconomic status: A machine learning approach. *Social Science & Medicine, 238*, 112486. https://doi.org/10.1016/j.socscimed.2019.112486

Daoud, A., & Dubhashi, D. (2023). Statistical modeling: The three cultures. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.89f6fe66

Daoud, A. (2018). Unifying studies of scarcity, abundance, and sufficiency. *Ecological Economics, 147*, 208–217. https://doi.org/10.1016/j.ecolecon.2018.01.019

Daoud, A., & Nandy, S. (2019). Implications of the politics of caste and class for child poverty in India. *Sociology of Development, 5*, 428–451. https://doi.org/10.1525/sod.2019.5.4.428

Deaton, A. (2015). The great escape: Health, wealth, and the origins of inequality, Reprint edition. (ed). Princeton University Press.

Decelle, A. (2022). Fundamental problems in statistical physics XIV: Lecture on machine learning. arXiv preprint arXiv:2202.05670. https://doi.org/10.48550/arXiv.2202.05670

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition. Presented at the 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

Doll, C. N. H., Muller, J.-P., & Morley, J. G. (2006). Mapping regional economic activity from nighttime light satellite imagery. *Ecological Economics, 57*, 75–92. https://doi.org/10.1016/j.ecolecon.2005.03.007

Drèze, J., & Sen, A. (2013). *An uncertain glory: India and its contradictions*. Penguin.

Ekstrand, S. (1996). Landsat TM-based forest damage assessment: Correction for topographic effects. *Photogrammetric Engineering and Remote Sensing, 62*, 151–162.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Gordon, D., & Nandy, S. (2012). Measuring child poverty and deprivation. In Z. Minujin, M. Alberto, & S. Nandy (Eds.), *Global child poverty and well-being: Measurement concepts policy and action* (pp. 57–101). Policy Press.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, Big Remotely Sensed Data: Tools, Applications and Experiences, 202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Halleröd, B., Rothstein, B., Daoud, A., & Nandy, S. (2013). Bad governance and poor children: A comparative analysis of government efficiency and severe child deprivation in 68 low-and middle-income countries. *World Development, 48*, 19–31.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Presented at the proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Head, A., Manguin, M., Tran, N., Blumenstock, J. E. (2017). Can human development be measured with satellite imagery? https://doi.org/10.1145/3136560.3136576

Henderson, J. V., Squires, T., Storeygard, A., & Weil, D. (2018). The Global distribution of economic activity: Nature, history, and the role of trade. *The Quarterly Journal of Economics, 133*, 357–406. https://doi.org/10.1093/qje/qjx030

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review, 102*, 994–1028. https://doi.org/10.1257/aer.102.2.994

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*, 790–794. https://doi.org/10.1126/science.aaf7894

Jerzak, C.T., Johansson, F., Daoud, A. (2023). Integrating earth observation data into causal inference: challenges and opportunities. arXiv preprint arXiv:2301.12985

Khatami, R., Mountrakis, G., & Stehman, S. V. (2016). A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment, 177*, 89–100. https://doi.org/10.1016/j.rse.2016.02.028

Kino, S., Hsu, Y.-T., Shiba, K., Chien, Y.-S., Mita, C., Kawachi, I., & Daoud, A. (2021). A scoping review on the use of machine learning in research on social determinants of health: Trends and research prospects. *SSM-Population Health, 15*, 100836. https://doi.org/10.1016/j.ssmph.2021.100836

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*, 84–90. https://doi.org/10.1145/3065386

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems.* Morgan-Kaufmann.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539

McBride, L., & Nichols, A. (2016). Retooling poverty targeting using out-of-sample validation and machine learning. *The World Bank Economic Review*. https://doi.org/10.1093/wber/lhw056

Nandy, S., Daoud, A., & Gordon, D. (2016). Examining the changing profile of undernutrition in the context of food price rises and greater inequality. *Social Science & Medicine, 149*, 153–163. https://doi.org/10.1016/j.socscimed.2015.11.036

Pandey, S.M., Agarwal, T., Krishnan, N.C. (2018). Multi-task deep learning for predicting poverty from satellite images. In *Thirty-second AAAI conference on artificial intelligence. Presented at the thirty-second AAAI conference on artificial intelligence*.

Papadakis, N. (2015). Optimal transport for image processing. In *Habilitation thesis*, Université de Bordeaux.

Raheja, S., Obaidat, M. S., Sadoun, B., Malik, S., Rani, A., Kumar, M., & Stephan, T. (2021). Modeling and simulation of urban air quality with a 2-phase assessment technique. *Simulation Modelling Practice and Theory, 109*, 102281. https://doi.org/10.1016/j.simpat.2021.102281

Raheja, S., Obaidat, M. S., Kumar, M., Sadoun, B., & Bhushan, S. (2022). A hybrid MCDM framework and simulation analysis for the assessment of worst polluted cities. *Simulation Modelling Practice and Theory, 118*, 102540. https://doi.org/10.1016/j.simpat.2022.102540

Randall, S., & Coast, E. (2015). Poverty in African households: The limits of survey and census representations. *The Journal of Development Studies, 51*, 162–177. https://doi.org/10.1080/00220388.2014.968135

Reddy, S. G., & Daoud, A. (2020). Entitlements and capabilities. In E. C. Martinetti, S. Osmani, & M. Qizilbash (Eds.), *The cambridge handbook of the capability approach.* Cambridge University Press.

Riano, D., Chuvieco, E., Salas, J., & Aguado, I. (2003). Assessment of different topographic corrections in landsat-TM data for mapping vegetation types (2003). *IEEE Transactions on Geoscience and Remote Sensing, 41*, 1056–1061. https://doi.org/10.1109/TGRS.2003.811693

Richter, R., Kellenberger, T., & Kaufmann, H. (2009). Comparison of topographic correction methods. *Remote Sensing, 1*, 184–196. https://doi.org/10.3390/rs1030184

Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., Recht, B., & Hsiang, S. (2021). A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications, 12*, 4392. https://doi.org/10.1038/s41467-021-24638-z

Samriya, J. K., Tiwari, R., Cheng, X., Singh, R. K., Shankar, A., & Kumar, M. (2022). Network intrusion detection using ACO-DNN model with DVFS based energy optimization in cloud framework. *Sustainable Computing: Informatics and Systems, 35*, 100746. https://doi.org/10.1016/j.suscom.2022.100746

Simonyan, K., Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [cs].

Singh, A., Obaidat, M. S., Singh, S., Aggarwal, A., Kaur, K., Sadoun, B., Kumar, M., & Hsiao, K.-F. (2022). A simulation model to reduce the fuel consumption through efficient road traffic modelling. *Simulation Modelling Practice and Theory, 121*, 102658. https://doi.org/10.1016/j.simpat.2022.102658

Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A. M., Hadiuzzaman, K. N., Lu, X., Wetter, E., Tatem, A.

J., & Bengtsson, L. (2017). Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface, 14*, 20160690. https://doi.org/10.1098/rsif.2016.0690

Subash, S. P., Kumar, R. R., & Aditya, K. S. (2018). Satellite data and machine learning tools for predicting poverty in rural India. *Agricultural Economics Reseaerch Review, 31*, 231. https://doi.org/10.5958/0974-0279.2018.00040.X

Subramanian, S. V., Ambade, M., Kumar, A., Chi, H., Joe, W., Rajpal, S., & Kim, R. (2023). Progress on sustainable development goal indicators in 707 districts of India: A quantitative mid-line assessment using the national family health surveys, 2016 and 2021. *The Lancet Regional Health-Southeast Asia*. https://doi.org/10.1016/j.lansea.2023.100155

Suraj, P.K., Gupta, A., Sharma, M., Paul, S.B., Banerjee, S. (2017). On monitoring development indicators using high resolution satellite images. arXiv:1712.02282 [cs, econ].

Sutton, P. C., Elvidge, C. D., & Ghosh, T. (2007). Estimation of gross domestic product at sub-national scales using Nighttime satellite imagery. *International Journal of Ecological Economics & Statistics, 8*, 5–21.

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data, 4*, 170004. https://doi.org/10.1038/sdata.2017.4

Thorat, A., Vanneman, R., Desai, S., & Dubey, A. (2017). Escaping and falling into poverty in India today. *World Development, 93*, 413–426. https://doi.org/10.1016/j.worlddev.2017.01.004

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., Restaino, R., & Wald, L. (2015). A Critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing, 53*, 2565–2586. https://doi.org/10.1109/TGRS.2014.2361734

Watmough, G. R., Atkinson, P. M., Saikia, A., & Hutton, C. W. (2016). Understanding the evidence base for poverty-environment relationships using remotely sensed satellite data: An example from Assam, India. *World Development, 78*, 188–203. https://doi.org/10.1016/j.worlddev.2015.10.031

Xie, M., Jean, N., Burke, M., Lobell, D., Ermon, S. (2015). Transfer learning from deep features for remote sensing and poverty mapping. arXiv:1510.00098 [cs].

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications, 11*, 2583. https://doi.org/10.1038/s41467-020-16185-w

Young, N. E., Anderson, R. S., Chignell, S. M., Vorster, A. G., Lawrence, R., & Evangelista, P. H. (2017). A survival guide to landsat preprocessing. *Ecology, 98*, 920–932. https://doi.org/10.1002/ecy.1730

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive survey on transfer learning. *Proceedings of the IEEE, 109*, 43–76. https://doi.org/10.1109/JPROC.2020.3004555

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.