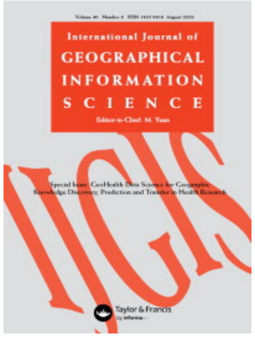

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: www.tandfonline.com/journals/tgis20

High-resolution mapping of demographic compositions based on remote sensing and social media data: a multimodal, multi-output modeling approach

Ge Qiu , Yuchen Li , Shaoqing Dai , Kun Qin , Zhanpeng Wang , Yaolin Liu , Jiping Liu & Peng Jia

To cite this article: Ge Qiu , Yuchen Li , Shaoqing Dai , Kun Qin , Zhanpeng Wang , Yaolin Liu , Jiping Liu & Peng Jia (27 Jul 2026): High-resolution mapping of demographic compositions based on remote sensing and social media data: a multimodal, multi-output modeling approach, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2026.2702485

To link to this article: https://doi.org/10.1080/13658816.2026.2702485

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2026.2702485

Taylor & Francis
by informa

RESEARCH ARTICLE

High-resolution mapping of demographic compositions based on remote sensing and social media data: a multimodal, multi-output modeling approach

Ge Qiu[a,b,c,d]†, Yuchen Li[e]†, Shaoqing Dai[a], Kun Qin[a,d], Zhanpeng Wang[a,d], Yaolin Liu[f], Jiping Liu[a,c] and Peng Jia[a,b,d]

[a]School of Resource and Environmental Sciences, Center for Metabolic and Panvascular Diseases, Wuhan University, Wuhan, China; [b]Medical Remote Sensing Information Research Institute (MRSIRI), Renmin Hospital of Wuhan University, Wuhan, China; [c]Chinese Academy of Surveying and Mapping, Beijing, China; [d]International Institute of Spatial Lifecourse Health (ISLE), Wuhan University, Wuhan, China; [e]School of Geography, University of Leeds, Leeds, UK; [f]Duke Kunshan University, Kunshan, China

ABSTRACT

Understanding fine-scale demographic compositions, such as sex and age structures, is fundamental to many fields, including public health, urban planning, and marketing. However, most existing fine-scale population products did not have spatial heterogeneities of demographic compositions within administrative units. This study, for the first time, developed a novel approach to estimate complete sex and age structures on 100×100 m grid cells, based on multimodal data including census, remote sensing imagery, and geotagged social media data. Key environmental variables derived from remote sensing imagery – such as artificial light intensity and vegetation index – and demographics-related variables extracted from social media texts were integrated using a multi-output random forest model, which, accounting for correlations among multiple demographic attributes, estimated the associations between the aforementioned variables and census-derived demographic compositions, including sex and age structures. The root mean square error (RMSE) of the final product was 6.18 for the sex ratio (the number of males per 100 females), and 1.89%, 3.82%, and 3.40% for the percentages over the total population of age groups 0–14, 15–59, and $\geq$60 years, respectively. This approach holds potential for detailed demographic modeling, which can be extended to other sociodemographic factors and benefit multiple disciplines.

KEY POLICY HIGHLIGHTS

- This study is the first to estimate complete sex and age structures at a 100-meter resolution.
- A multimodal and multi-output framework was developed to jointly model multiple demographic attributes using remote sensing imagery and social media texts.

ARTICLE HISTORY
Received 3 September 2024
Accepted 8 July 2026

KEYWORDS
Demographic composition; remote sensing; social media; natural language processing; multi-output model

CONTACT Jiping Liu liujp@casm.ac.cn; Peng Jia jiapengff@hotmail.com

Supplemental data for this article can be accessed online at https://doi.org/10.1080/13658816.2026.2702485.

†These authors contributed equally to this work.

- Contextualized features related to spatially detailed demographic compositions were extracted from social media texts.
- Social media data were found to be more influential in estimating sex structures than remote sensing data.

1. Introduction

Spatially fine-scale demographic compositions are prerequisites for various applications related to public health, planning, marketing, and fundamental elements for monitoring the progress of the Sustainable Development Goals (United Nations 2015, Szabo *et al*. 2020). Sex and age are among the most fundamental demographic characteristics (Drefahl *et al*. 2020, O'Connor *et al*. 2024). For instance, public health studies require spatially detailed sex and age structures, given that disease prevalence and mortality typically vary among different sex and age groups (Abdullah *et al*. 2022); urban planning requires such data for the precise allocation of public services and infrastructure, considering the different needs of various demographic groups (Calderón-Argelich *et al*. 2023); marketing relies on such data to tailor the user recommendations for shops catering to specific segments like women, children, and the elderly (Castronuovo *et al*. 2021). However, the widely used demographic composition data (e.g. census data) are usually represented by choropleth maps, illustrating demographic group percentages in the total population at coarse administrative levels (e.g. national, provincial), which are limited in capturing spatial heterogeneities within these units or integrating with other high-resolution gridded products (e.g. meteorological datasets) for broader use (Nilsen *et al*. 2021). Additionally, although existing gridded population products (e.g. WorldPop) provide total population counts at high resolutions, their sex- and age-specific information is directly derived from census data at coarse administrative levels, thereby neglecting fine-scale heterogeneities across grids (Pezzulo *et al*. 2017). Therefore, it is crucial to estimate sex and age structures within the fine-scale grids based on ancillary data.

Remote sensing data have been used to estimate demographic compositions, particularly age, as environmental factors (e.g. vegetation, terrain, and nightlight intensity) may reflect residential preferences of different age groups (Huang *et al*. 2022, Jeong and Lee 2026, Ju *et al*. 2025). However, there remains a lack of sufficient examination of the associations between various remote sensing data and the other demographic compositions, such as sex. This insufficiency may be because environmental factors captured by remote sensing are more easily linked to age-related behaviors and distributions, whereas other demographic attributes might require additional data sources for accurate modeling (Gebru *et al*. 2017, Steele *et al*. 2017, Suel *et al*. 2019). Newly emerging datasets from social media platforms such as Twitter and Sina Weibo hold significant potential for revealing demographic information due to the rich details they contain about user characteristics (Mislove *et al*. 2021, Fecht *et al*. 2020). Attributes such as usernames and language usage of Weibo posts could provide insights into users' demographic characteristics (Guimaraes *et al*. 2017, Nguyen *et al*. 2021), while the (x,y) spatial coordinates attached to these posts (i.e. geotags) enable the exploration of their spatial distributions. However, previous studies used social media texts aggregated at the administrative-unit level for estimation of demographic compositions, which did not fully leverage the fine-scale spatial information available in individual geotagged posts. Moreover, these studies extracted numeric features from unstructured texts by assigning fixed values to words, regardless of their meanings in different contexts (e.g. varying usage of a given word in different conditions across sexes and age groups). This neglect of contextual meanings limits the ability of the extracted features in capturing demographic differences (Ethayarajh 2019, Merchant *et al.* 2020, Zhu and Mao 2023). Therefore, more advanced methods are needed to retain fine-scale spatial information and extract context-aware numeric features from geotagged social media texts for high-resolution estimation of demographic compositions.

Machine learning models are widely used in high-resolution population estimation due to their ability to capture complex and nonlinear relationships. In particular, random forest has been one of the most frequently adopted models, owing to its strong robustness and interpretability (Ye *et al*. 2019, Zhang and Zhao 2024). However, its typical single-output design limits the ability to capture dependencies among correlated demographic variables such as multiple sex and age group percentages (Rahman *et al*. 2017, Cui *et al*. 2018). Neglecting these dependencies by training separate models can be time-consuming and potentially result in decreased predictive performance compared to utilizing a coherent multi-output model (Cui *et al*. 2018, Yang *et al*. 2023). For modeling two fully correlated sex groups, consolidating them into a single index such as the sex ratio (number of males per 100 females) is feasible. However, for age groups with more than two categories, this approach cannot capture all correlations, so multi-output learning could be a more effective solution.

To address these shortcomings, this study aimed to develop a novel approach to estimate sex ratio and % of different age groups at 100×100 m grid cells, using remote sensing and social media ancillary data. Numeric features were extracted from geotagged social media texts via advanced deep learning models and then processed through various spatial modeling techniques to generate spatial variables. These variables, combined with environmental variables from remote sensing imagery, were used as independent variables. The classic and multi-output random forest models were used to estimate associations of sex ratio and % of different age groups with these variables for demographic composition estimation. The major contributions of this study to the existing literature are as follows. First, for the first time, complete demographic composition maps – including both sex and age structures – were produced at a 100-m resolution, enabling the capture of fine-scale spatial heterogeneities that are overlooked in existing population products. Second, to enable such fine-scale demographic composition mapping, a novel feature extraction method was developed by integrating natural language processing with spatial modeling to derive demographics-related variables from geotagged social media texts, supplementing environmental variables obtained from remote sensing imagery. Third, dependencies across multiple interrelated demographic compositions were explicitly considered to improve estimation accuracy, using multi-output modeling approaches. The findings of this study would advance the implications and applicability of population products by providing spatially detailed demographic compositions, offering practical support for various fields depending on these data.

2. Related works

2.1. Gridded population datasets

Gridded population products provide standardized, georeferenced cells of population density, which are essential for effective planning and resource allocation across multiple fields. There have been several well-known global efforts to generate high-resolution gridded population data, such as Gridded Population of the World (GPW) (Balk and Yetman 2004), Global Rural-Urban Mapping Project (GRUMP) (Balk *et al*. 2006), Global Human Settlement Population Grid datasets (GHS-POP) (Freire *et al*. 2016), LandScan Global (Dobson *et al*. 2000), WorldPop (Tatem *et al*. 2013), High-Resolution Population Density Maps (HRPDM) (Tiecke *et al*. 2017), and Kontur (Kontur 2024). Although these products accurately reflect the total population counts at high resolutions (e.g. 1 km to 30 m grids), only GPW, WorldPop, and HRPDM provide sex and age information. For instance, GPW evenly allocates each census-derived demographic subgroup across grids within administrative units based on an area-weighting technique. WorldPop and HRPDM generate gridded total population estimates using ancillary data, which are subsequently split into demographic subgroups using demographic composition ratios at the administrative-unit levels (Pezzulo *et al*. 2017). As such, these products did not account for spatial heterogeneities of demographic compositions within administrative units, highlighting the need for generating demographic compositions at high resolutions.

2.2. Influential factors of demographic composition distributions

There have been several local studies to generate fine-scale demographic compositions. Alegana *et al*. (2015) employed the enhanced vegetation index (EVI), intensities of artificial light at night (ALAN), land cover, and accessibility to cities, to estimate the proportion of population aged 0–5 years across 1-km grids in Nigeria. Li *et al*. (2016) used land cover to estimate population densities of different age groups (i.e. aged 0–14, 15–64, and ≥65 years) onto 1-km gridded surfaces in Hungary. Jeong and Lee (2026) used more detailed environmental factors – such as urban land use, counts of different types of buildings, the distances to roads, and transportation stations – to estimate population densities of different age groups across 1-km grids in South Korea. Ju *et al*. (2025) further improved spatial resolution by producing 100-m gridded estimates of population densities for different age groups (i.e. aged 0–14, 15–59, 60–64, and ≥65 years) in China, using a broader set of environmental factors such as built-up area, building height, water bodies, elevation, and slope. These studies have provided accurate and fine-scale age structure data, offering valuable insights for various fields. However, they primarily rely on environmental factors for estimation, overlooking other data sources that could provide additional information related to demographic composition distributions. Additionally, these studies largely focus on mapping age structures while neglecting sex structures which are equally important for many fields such as public health and urban planning.

A limited number of studies have estimated sex structures at the administrative-unit level. For example, Montasser and Kifer (2017) used geotagged social media data to estimate sex and racial structures at the census-block level in the United States. They applied static embedding techniques (e.g. Word2Vec) to extract numeric features from a set of social media texts in each census block, and then estimated associations between the extracted features and population counts of different demographic groups. Although the utilization of such crowdsourced data necessitates careful consideration of representativeness bias (e.g. higher proportions of younger users) and privacy constraints (e.g. selective disclosure of user profiles), these studies have successfully demonstrated the potential of social media data for demographic composition estimation at a relatively fine scale (i.e. the census-block level). However, to achieve finer-scale estimates (e.g. 100-m grids), precise spatial coordinates of geotagged social media posts should be fully utilized, as aggregating data to coarser administrative units discards critical spatial details. Additionally, the static embedding techniques used in these studies assign the same feature vector to a word, even when it has distinct meanings in different contexts, which can result in less discriminative text representations for various demographic groups. For example, the verb 'crush' might be used by adolescent females to express romantic admiration toward celebrities (e.g. 'crushing on idols'), but by middle-aged males to denote competitive dominance in professional contexts (e.g. 'crushing competitors'). Therefore, it is essential to extract contextual features from social media texts using more advanced natural language processing methods, and employ spatial modeling techniques to fully utilize the fine-scale location information attached to individual posts for high-resolution demographic composition estimation.

Bidirectional encoder representations from transformers (BERT) is an advanced deep learning model for human language representation, which converts text inputs into semantically representative numeric features. It is pre-trained on a large amount of general and unlabeled corpora, and thus produces features that reflect general language usage (Vaswani *et al*. 2017, Devlin *et al*. 2019). Studies have fine-tuned BERT for specific tasks using labeled datasets, which updates parameters of BERT based on a limited amount of task-specific data without requiring full retraining, and thus enables the extracted features explicitly related to the task labels (Sun *et al*. 2019). For example, Nia *et al*. (2023) and Abdul-Mageed *et al*. (2019) fine-tuned BERT to identify users' demographic characteristics based on their posts published on Twitter. Although these studies employed fine-tuned BERT to extract demographically relevant features, they focused on classification tasks and did not incorporate geospatial information for population mapping. Geotags of social media posts, typically generated when users allow location access while posting, can provide precise (x,y) spatial coordinates for high-resolution mapping. Spatial modeling techniques such as inverse distance weighting, which interpolates attribute values from nearby points, and kernel density estimation, which identifies spatial concentrations of points, can be used to generate spatial variables related to the distributions of demographic compositions.

2.3. Modeling approaches for demographic composition estimation

Random forest was one of the most frequently used models for estimating demographic compositions. Compared to other machine learning models, such as XGBoost, random forest adopts a bagging-based training strategy to build multiple decision trees independently and average their predictions (Breiman 2001), which makes it more robust to noisy and small-sample data (e.g. census data) (Qiu *et al.* 2020). Its relatively simple model structure and fewer hyperparameters also make the model more interpretable and computationally efficient, thereby facilitating practical use in estimating demographic compositions (Zhao *et al.* 2021a). For example, Ju *et al.* (2025) applied separate random forest models to estimate population densities of different age groups across China, while Zhao *et al.* (2021b) conducted a similar study for Beijing. However, training separate models for each demographic group increases computational burden, and may overlook the correlations among demographic compositions (e.g. areas with higher proportions of working-age population may have lower proportions of elderly residents), which could in turn reduce estimation accuracy. For example, Cui *et al.* (2018) compared multi-output and single-output models for estimating healthcare resource utilization and found that the multi-output model, which considers correlations among dependent variables, achieved higher accuracy. Qin *et al.* (2024) also showed that the multi-output model outperformed single-output models in estimating concentrations of multiple air pollutants. Additionally, although deep learning models can inherently capture correlations among multiple outputs through shared network architectures and joint optimization, their training typically requires large amounts of labeled data at fine scales (e.g. census data at household or grid levels) (Zhuang *et al.* 2021, Jeong and Lee 2026), which are often inaccessible due to confidentiality policies (Jia *et al.* 2014). Therefore, multi-output random forest models that are specifically designed for estimating multiple correlated response variables, should be applied for more effective and accurate demographic composition estimation. However, to the best of the authors' knowledge, no existing study has employed multi-output models for demographic composition estimation.

3. Methods

3.1. Study area

Chengdu is the capital of Sichuan Province and one of the eight central cities in China. It is located between 102°54′–104°53′E and 30°05′–31°26′N, covering a total area of 14,335 km^2 and administering 261 townships (Figure 1). At the end of 2020, the population of the city stood at 20.94 million.

3.2. Datasets

The datasets used in this study included census, remote sensing, and geotagged social media data (Table 1). The latest Census 2020 at the township level (equivalent to level 4 of the Global Administrative Unit Layer) were acquired from the National Bureau of Statistics of China as the demographic composition reference. The corresponding

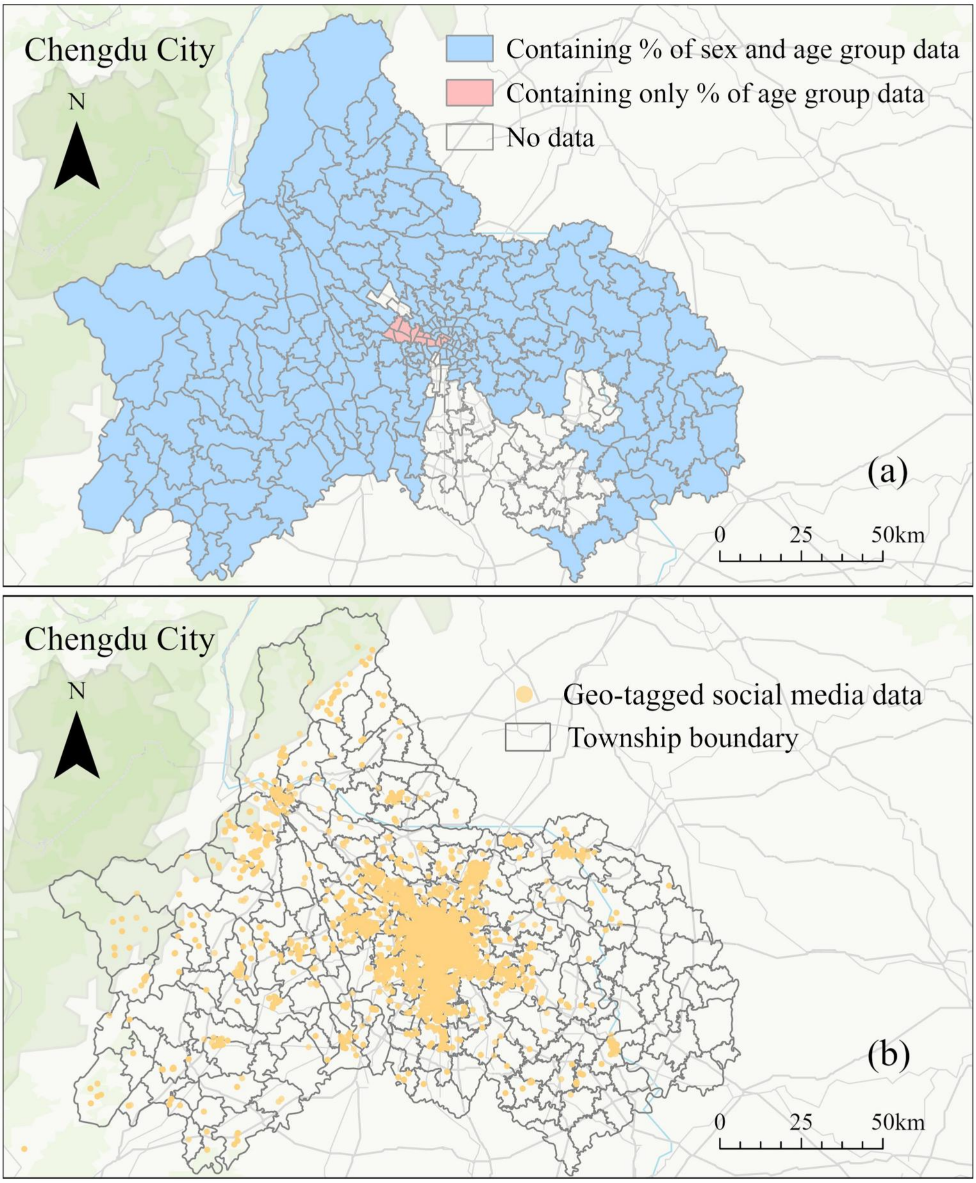

Figure 1. Demographic composition data coverage and the distribution of geotagged social media data in Chengdu City, China: (a) census data coverage by sex and age; (b) the distribution of geo-tagged social media data.

township-level administrative boundaries were obtained from the Administration of Surveying, Mapping and Geo-information. Among the 261 townships, 218 released both % of sex and age groups to the public (i.e. polygons in light blue in Figure 1), 12 provided only % of age groups (i.e. polygons in light pink in Figure 1), and the rest 31 have not yet disclosed detailed information of age or sex (i.e. polygons with transparency in Figure 1). The WorldPop age and sex structure dataset, one of the most

Table 1. Datasets used in this study.

Data	Format	Year	Description	Source
Census	Tables	2020	Township-level census containing % of males, % of females, % of age group 0–14 years, % of age group 15–59 years, % of age group ≥60 years	National Bureau of Statistics of China
Boundary maps	Polygon features	2020	Township-level administrative boundaries	Administration of Surveying, Mapping and Geo-information, China
WorldPop	Raster	2020	Gridded age and sex structures with a 100-m resolution	University of Southampton, UK
NPP-VIIRS ALAN	Raster	2020	Stable and cloud-free intensity of ALAN at a 500-m resolution	National Oceanic and Atmospheric Administration (NOAA), USA
MOD13Q1 EVI	Raster	2020	EVI dataset with a 250-m resolution	National Aeronautics and Space Administration (NASA), USA
SRTM Version 4	Raster	–	Elevation dataset at a 90-m resolution	Consultative Group for International Agricultural Research (CGIAR)
Geotagged Sina Weibos	Point features	2022	Sina Weibo posts with corresponding usernames, self-reported sex and age labels	Sina Com Technology Co., Ltd., China

ALAN, artificial light at night; EVI, enhanced vegetation index; MOD13Q1, Moderate-resolution Imaging Spectroradiometer Vegetation Indices; NPP-VIIRS, Suomi National Polar-Orbiting Partnership Visible Infrared Imaging Radiometer Suite; SRTM, Shuttle Radar Topography Mission.

accurate large-scale demographic products to date (Xu *et al*. 2021), was used to compare with the product of this study.

Remote sensing datasets that could reflect the distributions of demographic compositions were used as ancillary data. ALAN was obtained from Suomi National Polar-Orbiting Partnership Visible Infrared Imaging Radiometer Suite with a 500-m resolution (Elvidge *et al*. 2021). EVI was obtained from the Moderate-resolution Imaging Spectroradiometer with a 250-m resolution (Didan 2015). Elevation and slope were derived from the Shuttle Radar Topography Mission (SRTM) with a 90-m resolution (Farr and Kobrick 2000). Geotagged social media data were obtained from Sina Weibo which is the largest microblog platform in China and has about 511 million monthly active users according to its official report (Jiang *et al*. 2021). Geotagged Weibos are micro posts published by users with (x,y) spatial coordinates, which could be relevant to users' demographic characteristics (Gao *et al.* 2012). This study collected 31,983 geotagged Weibo posts across Chengdu City in August 2022, along with corresponding user details, including usernames, self-reported sex, and age. Usernames were available for all posts, while self-reported sex and age, provided optionally by users, were available for 31,681 and 17,174 posts, respectively. August, during which the Weibo data were collected, was carefully selected because it does not contain any national public holidays (e.g. Spring Festival in January or National Day in October) that could systematically affect

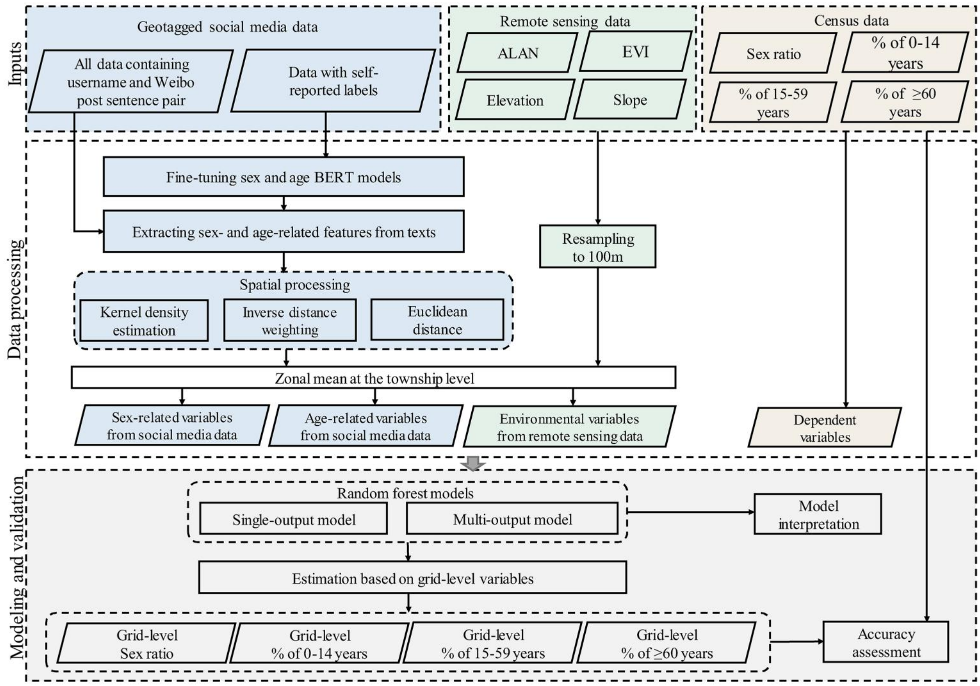

Figure 2. The flowchart of the general procedure in this study. ALAN, artificial light at night; BERT, bidirectional encoder representations from transformers; EVI, enhanced vegetation index.

posting behavior due to travel or large-scale social events. Additionally, to broadly examine the spatial representativeness of the Weibo data, the Pearson correlation coefficient between the number of Weibo posts and census-based population counts at the township level was calculated. The resulting coefficient of 0.54 indicates a moderate representativeness.

3.3. Data preparation with feature extraction from social media texts

Variables were extracted from the aforementioned datasets for demographic composition estimation (Figure 2). The finest demographic compositions at the township level – including sex ratio and % of age groups 0–14, 15–59, and $\geq$60 years – were extracted from census data as dependent variables. These age groupings follow the official demographic classification standard in China, which defines individuals aged 0–14 as children, 15–59 as the working-age population, and aged $\geq$60 as the elderly (Feng *et al*. 2019, National Bureau of Statistics 2021). Mean values of ALAN, EVI, elevation, and slope in every township were calculated as independent variables reflecting environmental characteristics.

BERT was fine-tuned to extract demographic features from social media texts (i.e. Weibo posts and usernames) (Figure 3). Specifically, two classification models were established, using self-reported sex (i.e. male and female) and age (i.e. 0–14, 15–59, and $\geq$60 years) as target labels, respectively. By updating parameters of BERT using small sets of labeled samples, the fine-tuned BERT models were able to associate text

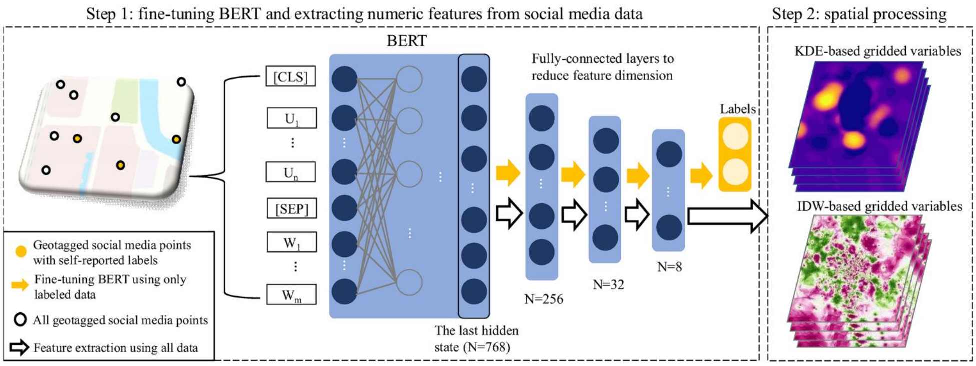

Figure 3. The procedure of feature extraction from social media texts and spatial processing. [CLS] is a special symbol to inform the start of text, and [SEP] is a separator dividing username (i.e. U_i) and Weibo post (i.e. W_j). BERT, bidirectional encoder representations from transformers; IDW, inverse distance weighting; KDE, kernel density estimation.

patterns with demographic categories and thereby generate features explicitly reflecting demographic information embedded in the text. To adapt these features for downstream spatial modeling, each classification model was further extended with three fully connected linear layers, reducing the original 768-dimensional BERT output to 8. The dimensionality of 8 was chosen as a trade-off between representation capacity and computational efficiency. Both models used 70% self-reported labels for training, 15% for validation, and 15% for testing. The Adam optimizer with weight decay was applied, and the learning rates were linearly decreased from 2×10^{-6} and 3×10^{-6} to 0 for sex and age classifications, respectively. The cross-entropy loss function was used for both models, with class weights adjusted based on label sample sizes to avoid class imbalance. The models were trained for 3 epochs using a batch size of 32, as recommended in previous studies (Devlin *et al*. 2019). After that, all pairs of Weibo posts and usernames were input into the fine-tuned BERT models to extract demographic features, with one model focusing on sex-related features and the other on age-related features.

The features extracted from geotagged posts were spatially modeled using kernel density estimation (KDE) and inverse distance weighting (IDW) to produce continuous gridded surfaces representing their spatial distribution. KDE-based gridded surfaces were characterized by high values indicating concentrated clusters of features with elevated intensity, which represent the estimated probability density of features and potentially reflect the spatial distribution of demographic compositions (Gatrell *et al*. 1996, Yang *et al*. 2019, Ye *et al*. 2019). IDW is a standard interpolation method for estimating values at unsampled locations based on distance-weighted averaging of nearby observed values (Lu and Wong 2008). The bandwidth of KDE and the power of IDW were set to 10 km and 1, respectively, to balance the representation of local spatial variation and the generation of sufficiently smooth continuous surfaces across the study area. This rationale is supported by previous studies applying KDE and IDW in modeling human activities and population-related variables (Asal 2012, Wu *et al*. 2018, Miao *et al*. 2021). Additionally, the spatial distribution of geotagged Weibo posts themselves – regardless of extracted feature values – may partially reflect demographic composition distribution. For example, younger people tend to be more active on social media and publish more geotagged posts. As such, the Euclidean distance to the nearest geotagged post was calculated from the center of each grid. This approach assumes that locations closer to geotagged posts could be more likely to contain populations sharing similar demographic characteristics, according to the first law of geography (Tobler 1970, Bakillah *et al*. 2014). Similarly, geotagged Weibo posts with self-reported sex and age labels were divided into different demographic groups (i.e. male and female for sex, and 0–14, 15–59, and $\geq$60 years for age), and the distance to the nearest geotagged post of each group was calculated. Mean values of each of the abovementioned gridded surfaces in every township were calculated as demographics-related independent variables (i.e. sets of sex-related and age-related variables).

3.4. Multi-output random forest modeling

The associations between independent variables and multiple correlated demographic attributes were estimated through our multi-output modeling approach. For the fully correlated sex attributes (i.e. % of males and females), this study consolidated them into a single index of sex ratio (i.e. the number of males per 100 females) to explicitly consider their dependence. A classic random forest model was then employed to estimate the associations of the sex ratio and environmental and sex-related independent variables. For the % of three different age groups, which are correlated but not fully dependent, a coherent multi-output random forest model was established to simultaneously estimate their associations with environmental and age-related independent variables. Extended from the classic random forest (Breiman 2001), the multi-output random forest employed a Mahalanobis splitting rule to construct regression trees (Ishwaran *et al*. 2021). This splitting rule incorporates the covariance matrix of multiple outputs, enabling the model to account for their joint variation and interdependencies when determining optimal splits (Segal and Xiao 2011, Tang and Ishwaran 2017). By explicitly considering the relationships among the multiple age groups, this approach may improve predictive performance and facilitate a more comprehensive interpretation of variable contributions (Evgeniou and Pontil 2004, Alakus *et al*. 2023). The general structure of the two models is shown as follows:

$$SR_i = f_{rf}(ENV_i, SEXR_i) \tag{1}$$

$$\begin{bmatrix} AG1_i \\ AG2_i \\ AG3_i \end{bmatrix} = f_{mrf}(ENV_i, AGER_i) \tag{2}$$

where SR_i denotes sex ratio in the township i; ENV_i denotes environmental independent variables from remote sensing data; $SEXR_i$ denotes sex-related independent variables from social media data; $f_{rf}()$ denotes the established classic random forest model; $AG1_i$, $AG2_i$, and $AG3_i$ denote % of age groups 0–14, 15–59, and $\geq$60 years in the township i, respectively; $AGER_i$ denotes age-related independent variables from social media data; $f_{mrf}()$ denotes the established multi-output random forest model. After establishing the associations based on the two models, values of independent variables within each 100-m grid were used to estimate demographic compositions at the grid level.

Variable importance measured by mean decrease accuracy (MDA) and partial dependence plots were used to visualize the established associations (Strobl *et al.* 2007, Scarpone *et al.* 2020). A locally weighted scatterplot smoothing function with a smooth span of 0.3 was employed to obtain clearer results of the partial dependence plots. The abovementioned procedures were implemented in R (version 4.1.3), with the sex ratio estimated using the randomForest package (Liaw and Wiener 2002) and the % of different age groups estimated using the randomForestSRC package (Ishwaran *et al.* 2021). The hyperparameter settings for both models are provided in Table S1.

3.5. Model comparison and accuracy assessment

To highlight the effectiveness of our modeling approach, several models frequently used for population mapping were established for comparison, including eXtreme Gradient Boosting (XGBoost), support vector machine (SVM), and feed-forward neural networks (FNN), all of which estimated demographic attributes separately without considering dependencies across them. In contrast, the multi-output random forest model using all demographic attributes as dependent variables (MRF-ALL) was also established for comparison, which imposed more constraints (i.e. sex categories) to capture potential dependencies among all demographic attributes and leveraged a broader set of independent variables, including both sex- and age-related variables, for more complex joint modeling. The abovementioned models were implemented in R (version 4.1.3), with XGBoost using the xgboost package (Chen *et al.* 2022), SVM using the e1071 package (Meyer *et al.* 2023), FNN using the nnet package (Venables and Ripley 2002), and MRF-ALL using the randomForestSRC package (Ishwaran *et al.* 2021). Moreover, to demonstrate the effectiveness of our feature extraction approach, this study extracted features using the original BERT and long short-term memory (LSTM) models (Hochreiter and Schmidhuber 1997), and compared the resulting demographic composition estimates with the features generated by our fine-tuned BERT models. The original BERT model reflected general language usage patterns without updating its parameters. The LSTM model captured only sequential patterns within texts, without the capacity to model global dependencies across all words as achieved by BERT. To ensure fair ablation experiments, the features extracted by the original BERT and LSTM models experienced the same spatial processing procedures and modeling method for estimating demographic compositions as those applied in our approach. Detailed settings of the LSTM model can be found in the Supplementary Material. Additionally, to assess whether the demographics-related variables generated from the social media data contribute to demographic composition estimation, a single-modality model using only remote sensing variables (RS-only) was also constructed for comparison.

The relative accuracy of gridded demographic compositions generated by the aforementioned models was compared. The mean values of grid-level sex ratio, % of age groups 0–14, 15–59, and $\geq$60 years in every township were calculated and compared with the census data. The WorldPop age and sex structure dataset was also included in the comparison. Root mean square error (RMSE) and mean absolute error (MAE) were used to quantify the differences between these products and the census data:

$$\text{RMSE} = \sqrt{\sum \left(D_{ij}^{estimated} - D_{ij}^{observed}\right)^2 / N} \tag{3}$$

$$\text{MAE} = \frac{1}{N} \sum \left| D_{ij}^{estimated} - D_{ij}^{observed} \right| \tag{4}$$

where $D_{ij}^{estimated}$ and $D_{ij}^{observed}$ denote estimated and census values of demographic attribute i (e.g. sex ratio) in township j, respectively; N denotes the number of townships included in the accuracy assessment. In addition, the error reduction percentages (ERP) relative to WorldPop were also calculated for the models developed in this study:

$$\text{ERP} = \frac{RMSE_{WorldPop,i} - RMSE_{model,i}}{RMSE_{WorldPop,i}} \times 100\% \tag{5}$$

where $RMSE_{WorldPop,i}$ and $RMSE_{model,i}$ denote the RMSE for demographic attribute i calculated according to Equation (3), for the WorldPop and the gridded demographic compositions generated by the models developed in this study, respectively.

3.6. Robustness analysis

The geotagged social media data used in this study may include highly active accounts (e.g. corporate accounts), which may distort the resulting demographic composition estimates. To quantify their potential impact, a robustness analysis was conducted by comparing demographic composition products generated with and without posts from such accounts. The highly active accounts were defined as either verified corporate accounts or accounts whose post count exceeded a statistical outlier threshold. The threshold was set at the third quartile of post counts across all accounts plus 1.5 times the interquartile range, following Tukey's fences method for outlier detection (Tukey 1977, Qiu *et al.* 2020). Moreover, the geotagged Weibo data used in this study were collected in August 2022, providing a representative snapshot, though user behaviors may vary over time (e.g. across months and years). Such volatility in user behaviors may introduce variations in posting frequency, spatial distribution, and demographic representativeness of Weibo posts. To assess whether our modeling approach was sensitive to those potential variations, another robustness analysis was conducted. A total of 20% and 40% of Weibo posts were removed or duplicated, representing moderate and relatively large variations. The selection of posts for removal or duplication was fully random to avoid introducing systematic bias and to simulate plausible variations in the composition and distribution of Weibo data. The per-grid impact of the abovementioned perturbations (i.e. highly active accounts and volatility of user behaviors) on demographic composition estimates was quantified as follows:

$$Standardized\ impact = \left| \frac{D'_{ij} - D_{ij}}{SD_i} \right| \times 100\% \tag{6}$$

where D_{ij} and D'_{ij} denote the demographic attribute i estimated by the original model and the model excluding highly active accounts or considering volatility of user behaviors at grid j, respectively; SD_i denotes the standard deviation of demographic attribute *i* estimated by the original model across the study area. The standardized impacts below 100% indicate that the differences caused by excluding highly active accounts or considering volatility of user behaviors did not exceed the overall variability of the original estimates, suggesting that our modeling approach was relatively robust and not overly sensitive to these perturbations (Qiu *et al*. 2025).

4. Results

4.1. Characteristics of variables for demographic composition estimation

The average accuracy and Matthews correlation coefficient for the fine-tuned BERT models were 75.01% and 0.46 for sex classification, and 71.76% and 0.26 for the age group classification on the testing datasets. Demographic features were extracted based on the two fine-tuned models and spatially processed as independent variables (Table 2).

4.2. Comparisons of different modeling approaches

The relative accuracy of demographic composition products generated by different modeling approaches was assessed by comparing them with census data at the township level (Table 3). Results suggest that our modeling approach achieved the highest accuracy for estimating sex ratio, % of age groups 15–59 and ≥60 years among the evaluated methods, including XGBoost, SVM, FNN, and MRF-ALL. XGBoost, however, achieved better accuracy for estimating the % of age group 0–14 years. Moreover, using fine-tuned BERT models for feature extraction resulted in demographic composition products with the highest accuracy across all demographic attributes, compared to those generated using features extracted by other methods, including the original BERT and LSTM models. In addition, compared with the RS-only model that relied solely on remote sensing variables, our approach incorporating both remote sensing and social media variables achieved higher accuracy across all demographic attributes, demonstrating the added value of social media data for demographic composition estimation. Additionally, the Out-of-Bag RMSE values of our approach were 7.20 for sex ratio, and 2.03%, 3.83%, and 3.58% for % of age groups 0–14, 15–59, and ≥60 years, respectively. These values were comparable to the corresponding external validation RMSE values, suggesting no severe overfitting.

The maps of gridded sex ratio and % of different age groups generated by our modeling approach are shown in Figure 4. The sex ratio (i.e. the number of males per 100 females) was found to be lower in the center of the city, but higher in rural areas. The % of age group 0–14 years was lower in the center of the city, but relatively higher in suburban and eastern rural areas. The spatial distributions of % of age groups 15–59 and ≥60 years were found to be almost inverse, with higher % of age group 15–59 years and lower % of age group ≥60 years in urban areas.

Table 2. Basic characteristics of variables at the township level.

Variables	Mean ± standard deviation	
	Samples for sex ratio modeling (N = 218)	Samples for % of age groups modeling (N = 230)
Dependent variables		
Sex ratio (the number of males per 100 females)	102.93 ± 7.98	
% of age group 0–14 years		12.87 ± 2.39
% of age group 15–59 years		64.57 ± 7.56
% of age group ≥60 years		22.56 ± 7.30
Environmental independent variables from remote sensing data		
ALAN (nW/cm^2/sr)	12.30 ± 15.04	13.52 ± 15.75
EVI	0.36 ± 0.10	0.35 ± 0.10
Elevation (m)	582.97 ± 261.79	579.06 ± 255.39
Slope (°)	4.28 ± 5.50	4.11 ± 5.40
Demographics-related independent variables from social media data		
Distance to the nearest unlabeled geotagged post (km)	3.50 ± 4.89	3.34 ± 4.81
Distance to the nearest male-labeled geotagged post (km)	4.08 ± 5.22	
Distance to the nearest female-labeled geotagged post (km)	3.62 ± 4.89	
Distance to the nearest geotagged post labeled 0–14 years (km)		6.22 ± 6.50
Distance to the nearest geotagged post labeled 15–59 years (km)		3.60 ± 4.89
Distance to the nearest geotagged post labeled ≥60 years (km)		8.07 ± 8.54
KDE-based sex-related deep feature 1	0.05 ± 0.11	
KDE-based sex-related deep feature 2	−0.004 ± 0.01	
KDE-based sex-related deep feature 3	0.04 ± 0.09	
KDE-based sex-related deep feature 4	0.08 ± 0.17	
KDE-based sex-related deep feature 5	0.02 ± 0.05	
KDE-based sex-related deep feature 6	0.11 ± 0.23	
KDE-based sex-related deep feature 7	−0.04 ± 0.10	
KDE-based sex-related deep feature 8	0.08 ± 0.17	
IDW-based sex-related deep feature 1	0.32 ± 0.28	
IDW-based sex-related deep feature 2	0.01 ± 0.12	
IDW-based sex-related deep feature 3	0.18 ± 0.40	
IDW-based sex-related deep feature 4	0.54 ± 0.35	
IDW-based sex-related deep feature 5	0.14 ± 0.15	
IDW-based sex-related deep feature 6	0.60 ± 0.67	
IDW-based sex-related deep feature 7	−0.21 ± 0.36	
IDW-based sex-related deep feature 8	0.45 ± 0.46	
KDE-based age-related deep feature 1		0.13 ± 0.24
KDE-based age-related deep feature 2		−0.06 ± 0.11
KDE-based age-related deep feature 3		−0.23 ± 0.43
KDE-based age-related deep feature 4		0.23 ± 0.43
KDE-based age-related deep feature 5		0.07 ± 0.14
KDE-based age-related deep feature 6		−0.06 ± 0.12
KDE-based age-related deep feature 7		−0.12 ± 0.22
KDE-based age-related deep feature 8		−0.15 ± 0.28
IDW-based age-related deep feature 1		0.85 ± 0.25
IDW-based age-related deep feature 2		−0.45 ± 0.04
IDW-based age-related deep feature 3		−1.61 ± 0.21
IDW-based age-related deep feature 4		1.63 ± 0.21
IDW-based age-related deep feature 5		0.52 ± 0.09
IDW-based age-related deep feature 6		−0.46 ± 0.08
IDW-based age-related deep feature 7		−0.85 ± 0.14
IDW-based age-related deep feature 8		−1.07 ± 0.14

ALAN, artificial light at night; EVI, enhanced vegetation index; IDW, inverse distance weighting; KDE, kernel density estimation.

Table 3. Relative accuracy of demographic composition products generated by different methods.

		Demographic compositions			
Methods	Accuracy metrices	Sex ratio	% of age group 0–14 years	% of age group 15–59 years	% of age group ≥60 years
This study	RMSE	6.18	1.89%	3.82%	3.40%
	MAE	3.86	1.44%	2.54%	2.64%
	ERP	25.15%	29.92%	33.95%	41.36%
XGBoost	RMSE	6.68	1.78%	3.86%	3.42%
	MAE	4.13	1.36%	2.61%	2.70%
	ERP	18.97%	33.70%	33.34%	40.90%
SVM	RMSE	7.96	2.39%	7.68%	7.29%
	MAE	5.51	1.89%	6.28%	6.17%
	ERP	3.54%	11.19%	−32.59%	−25.84%
FNN	RMSE	7.53	2.32%	5.58%	5.51%
	MAE	4.94	1.86%	4.10%	4.48%
	ERP	8.67%	13.78%	3.63%	4.91%
MRF-All	RMSE	8.10	1.97%	4.11%	3.73%
	MAE	5.74	1.50%	2.89%	2.92%
	ERP	1.82%	26.93%	29.08%	35.62%
Original BERT	RMSE	6.34	1.89%	3.87%	3.48%
	MAE	4.10	1.45%	2.56%	2.70%
	ERP	23.19%	29.59%	33.22%	39.93%
LSTM	RMSE	6.48	1.93%	3.90%	3.47%
	MAE	4.15	1.47%	2.56%	2.72%
	ERP	21.51%	28.38%	32.62%	40.01%
RS-only	RMSE	6.93	1.91%	4.12%	3.70%
	MAE	4.40	1.46%	2.64%	2.88%
	ERP	16.00%	29.00%	28.84%	36.10%
WorldPop	RMSE	8.25	2.69%	5.79%	5.79%
	MAE	5.86	2.02%	4.54%	4.92%
	ERP	–	–	–	–

LSTM denotes feature extraction using the long short-term memory model, and Original BERT denotes feature extraction using the original bidirectional encoder representations from transformers model. ERP, error reduction percentage relative to WorldPop; FNN, feed-forward neural networks; MRF-ALL, multi-output random forest using all demographic attributes as dependent variables; MAE, mean absolute error; RMSE, root mean square error; RS-only, the single-modality model using only remote sensing variables; SVM, support vector machine; XGBoost, extreme gradient boosting.

4.3. Evaluations of model robustness

The geotagged posts published by active users and corporate accounts were found to have a minimum impact on the resulting demographic composition estimates (Figure 5a,b). More than 90% of grids exhibit standardized impacts below 100%, indicating that the differences introduced by those accounts were generally smaller than the inherent variability of the original estimates. Potential variations in social media user behaviors lead to slightly larger impacts on demographic composition estimates, yet about 80% of grids still exhibit standardized impacts below 100% (Figure 5c–f). This suggests that our modeling approach was relatively robust to these perturbations.

4.4. Variable importance for demographic compositions modeling

For estimating sex ratio, the distances to the nearest female- and male-labeled geotagged posts, ALAN, the KDE-based sex-related deep feature 2, and the distance to the nearest unlabeled geotagged post were the most important variables, according to the rank of MDA (Figure 6). Demographics-related variables from social media data contributed approximately 80% to the total variable importance, and the remaining

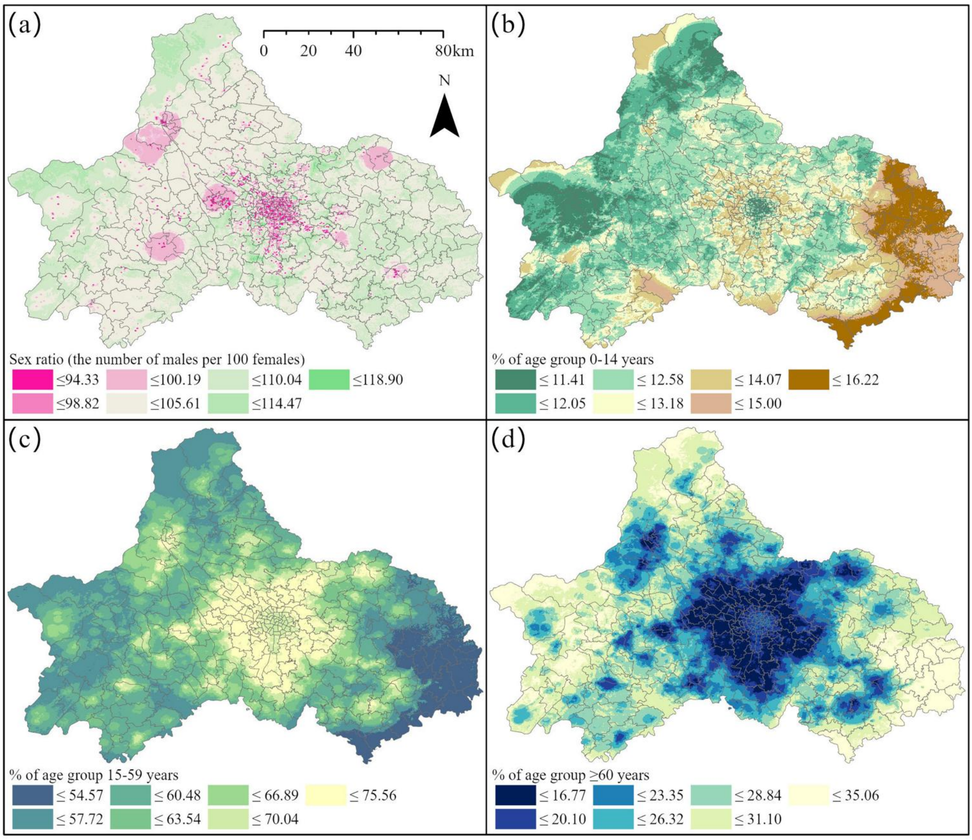

Figure 4. Grid-level maps of (a) sex ratio, % of age groups (b) 0–14, (c) 15–59, and (d) ≥60 years in Chengdu City in 2020.

20% came from environmental variables from remote sensing data (Figure 7). For estimating % of age group 0–14 years, the most important variables included the distance to the nearest geotagged post labeled 0–14 years, elevation, EVI, the distance to the nearest geotagged post labeled ≥60 years, and ALAN. Social media variables contributed about 61% to the total variable importance. For estimating % of age group 15–59 years, the most important variables included ALAN, the distance to the nearest unlabeled geotagged post, the distance to the nearest geotagged post labeled 0–14 years, KDE-based age-related deep feature 5, and deep feature 2. Social media variables accounted for about 56% of the overall variable importance. When estimating % of age group ≥60 years, ALAN, distance to the nearest unlabeled geotagged post, KDE-based age-related deep feature 5, deep feature 2, and the distance to the nearest geotagged post labeled 0–14 years were the most crucial variables. Social media variables comprised roughly 50% of the overall variable importance.

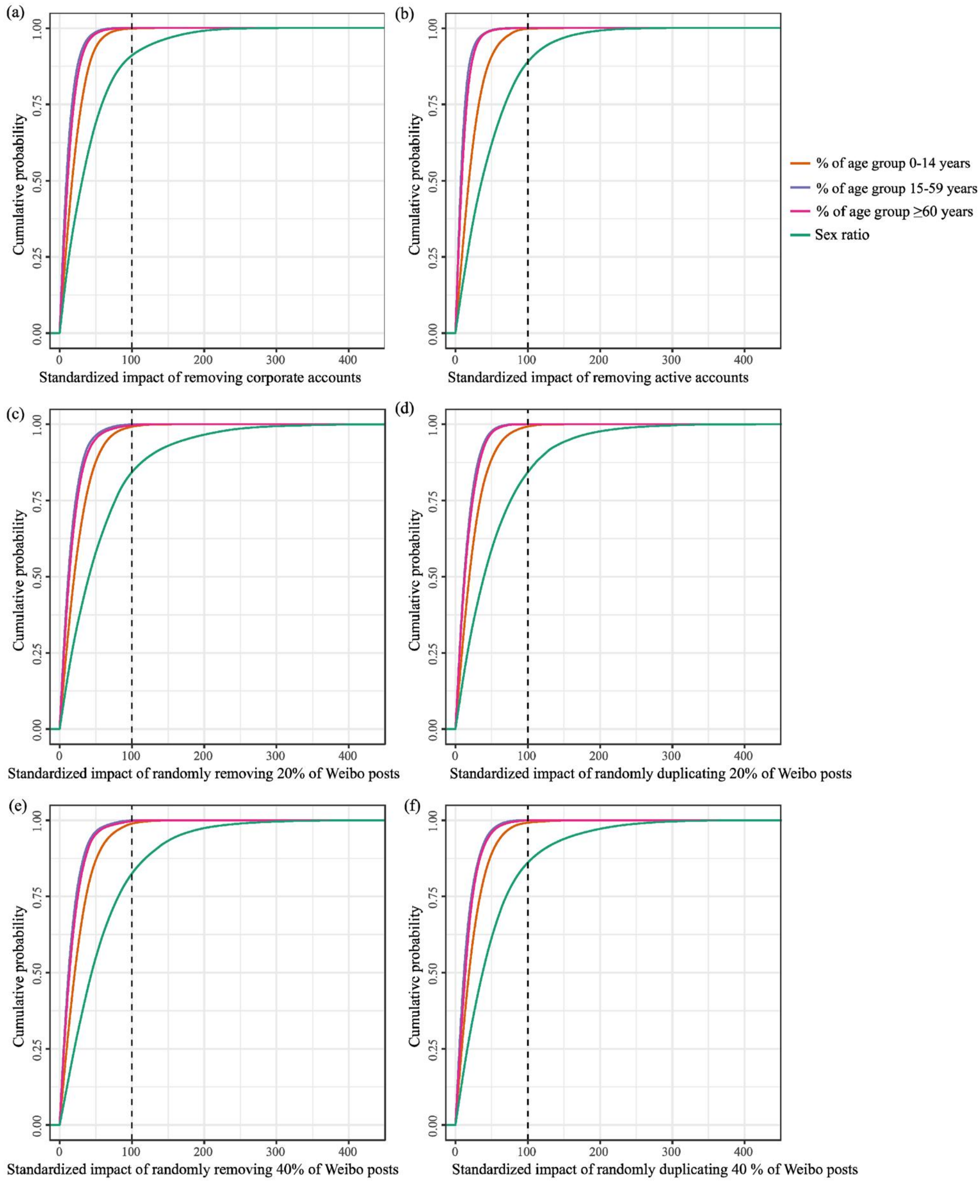

Figure 5. Cumulative probability curves of per-grid standardized impacts of different types of perturbations.

4.5. Associations between independent variables and demographic compositions

The association between sex ratio and elevation followed a non-linear trend (Figure 8). Initially, the sex ratio decreased rapidly to a minimum as elevation ascended to approximately 750 m, after which it exhibited a slight upward trend. The association between sex ratio and slope was generally positive, as higher sex ratios were associated with steeper slopes. The associations of sex ratio with ALAN and EVI were

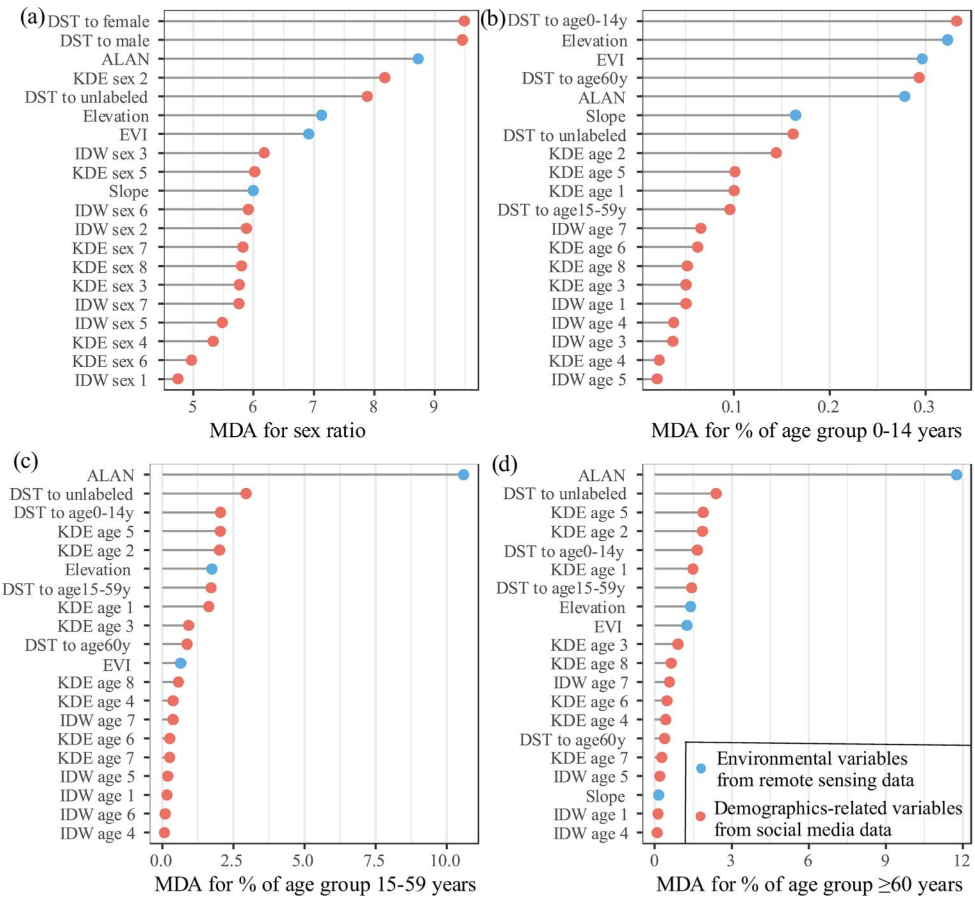

Figure 6. The twenty most important variables ranked by mean decrease accuracy (MDA). 'DST to *x*' denotes the distance to the nearest geotagged post labeled *x*, while 'IDW age/sex *x*' and 'KDE age/sex *x*' denote age- or sex-related deep feature *x*, spatially modeled using inverse distance weighting and kernel density estimation, respectively. ALAN, artificial light at night; EVI, enhanced vegetation index.

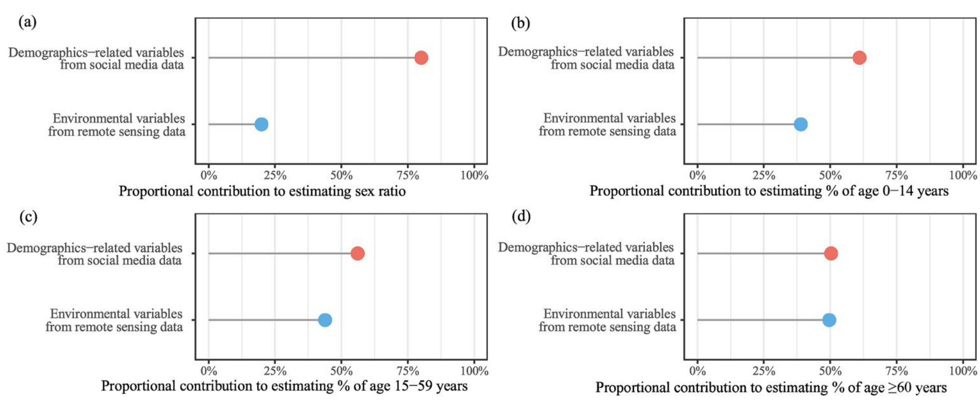

Figure 7. Proportional contributions of social media and remote sensing variables to demographic composition estimation.

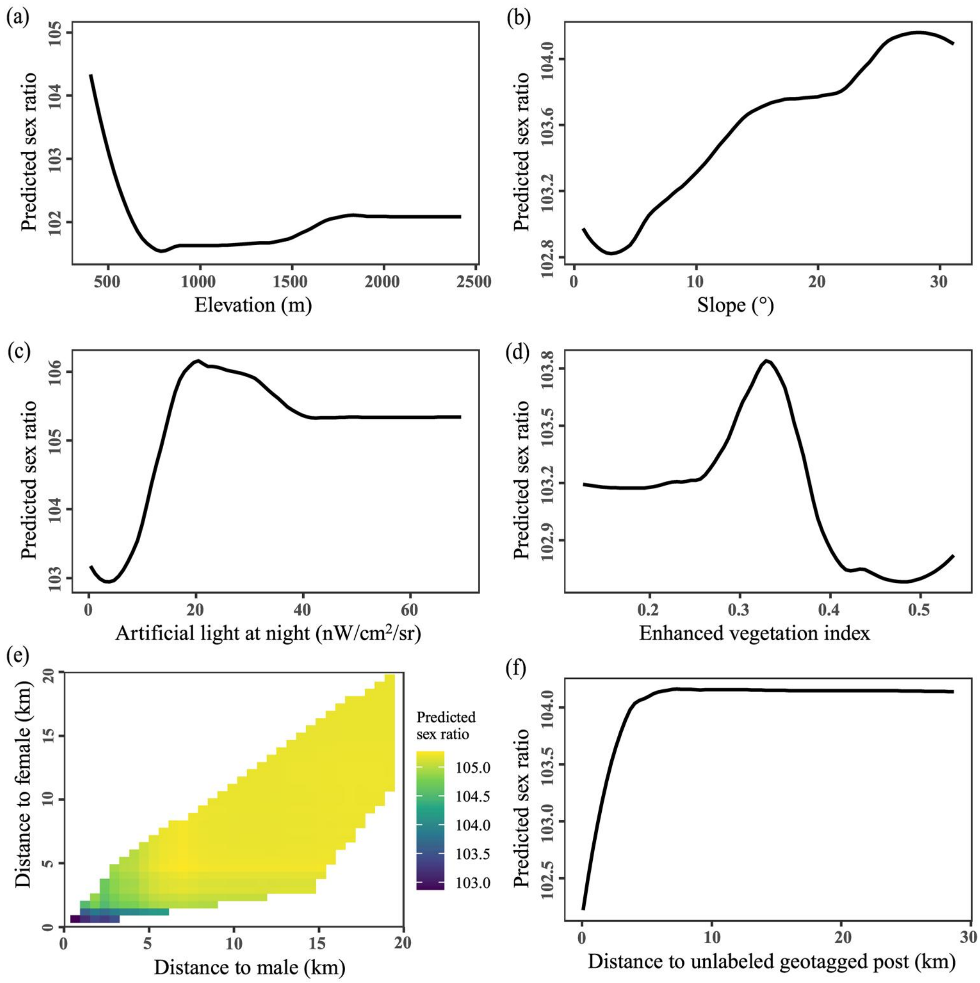

Figure 8. Associations between sex ratio and key independent variables. 'Distance to *x*' denotes the distance to the nearest geotagged post labeled *x*.

generally inverse V-shaped curves. The sex ratios increased to the peaks and then decreased, as ALAN increased to around 20 nW/cm^2/sr and EVI increased to approximately 0.35. For the two most important variables of the distances to the nearest female- and male-labeled geotagged posts, positive associations were found between the sex ratio and each of them, with a sharper increase trend as the distance to female rose. The sex ratio was positively associated with the distance to the nearest unlabeled geotagged post, with a rapid increase from 0 to around 5 km and then almost flattening. The associations between sex ratio and the KDE- and IDW-based sex-related deep features are shown in Figure S1. The semantic interpretations of these deep features are summarized in Table S2, with detailed derivation procedures provided in the Supplementary Material.

The % of age group 0–14 years was negatively associated with elevation, with a rapid decrease around 700 m and then almost flattening (Figure 9). The association between % of age group 0–14 years and slope exhibits an inverse V-shaped curve, and

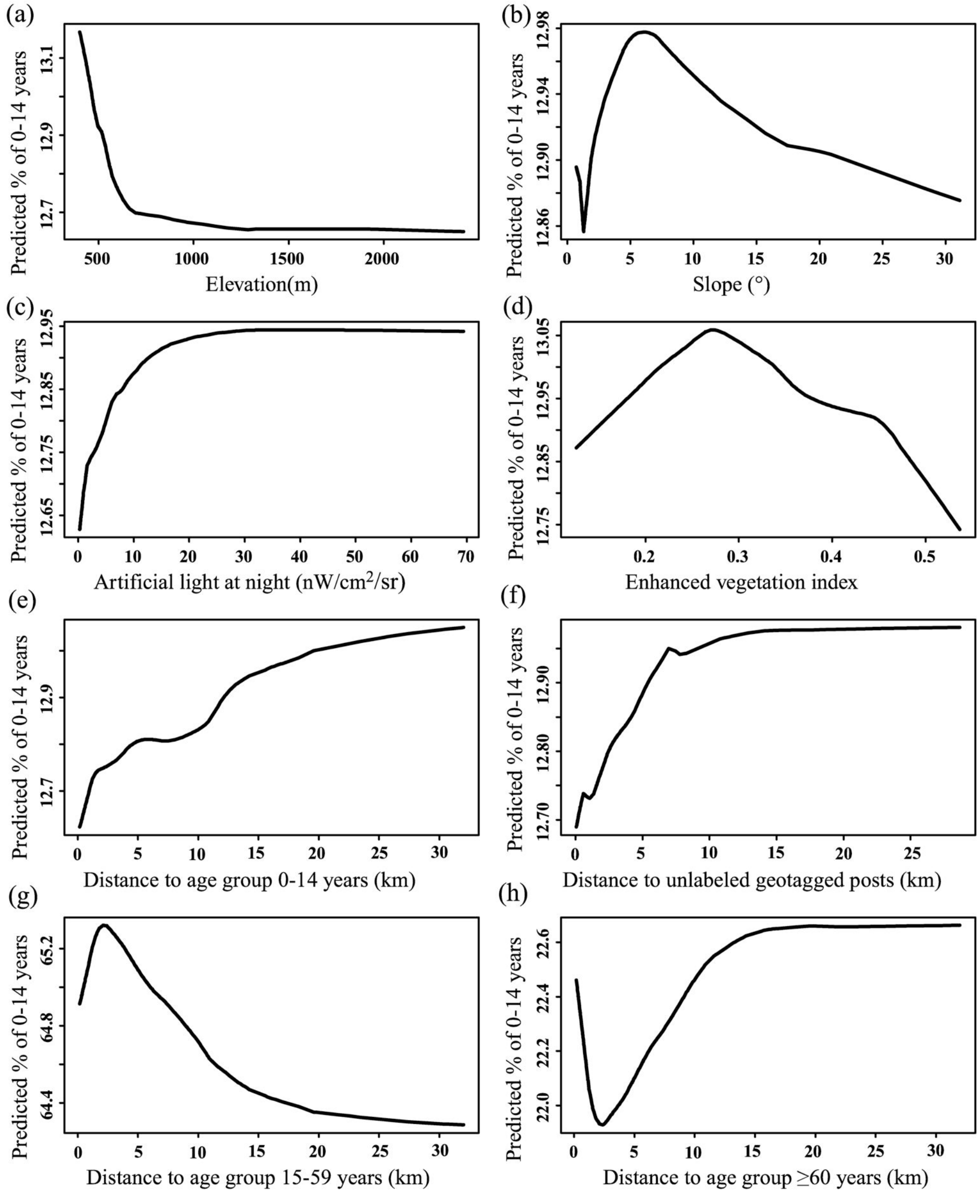

Figure 9. Associations between % of age group 0–14 years and key independent variables. 'Distance to *x*' denotes the distance to the nearest geotagged post labeled *x*.

the turning point is approximately above 5 degrees. As ALAN increased from 0 to 20 nW/cm^2/sr, the % of age group 0–14 years increased rapidly, which then stabilized. The association between % of age group 0–14 years and EVI exhibits a non-linear pattern. First, the % of age group 0–14 years increases as EVI rises to below 0.3, and then drops to a minimum. The % of age group 0–14 years was positively associated with both the distance to the nearest geotagged post labeled 0–14 years and the distance to the nearest unlabeled geotagged post. The associations between the % of age

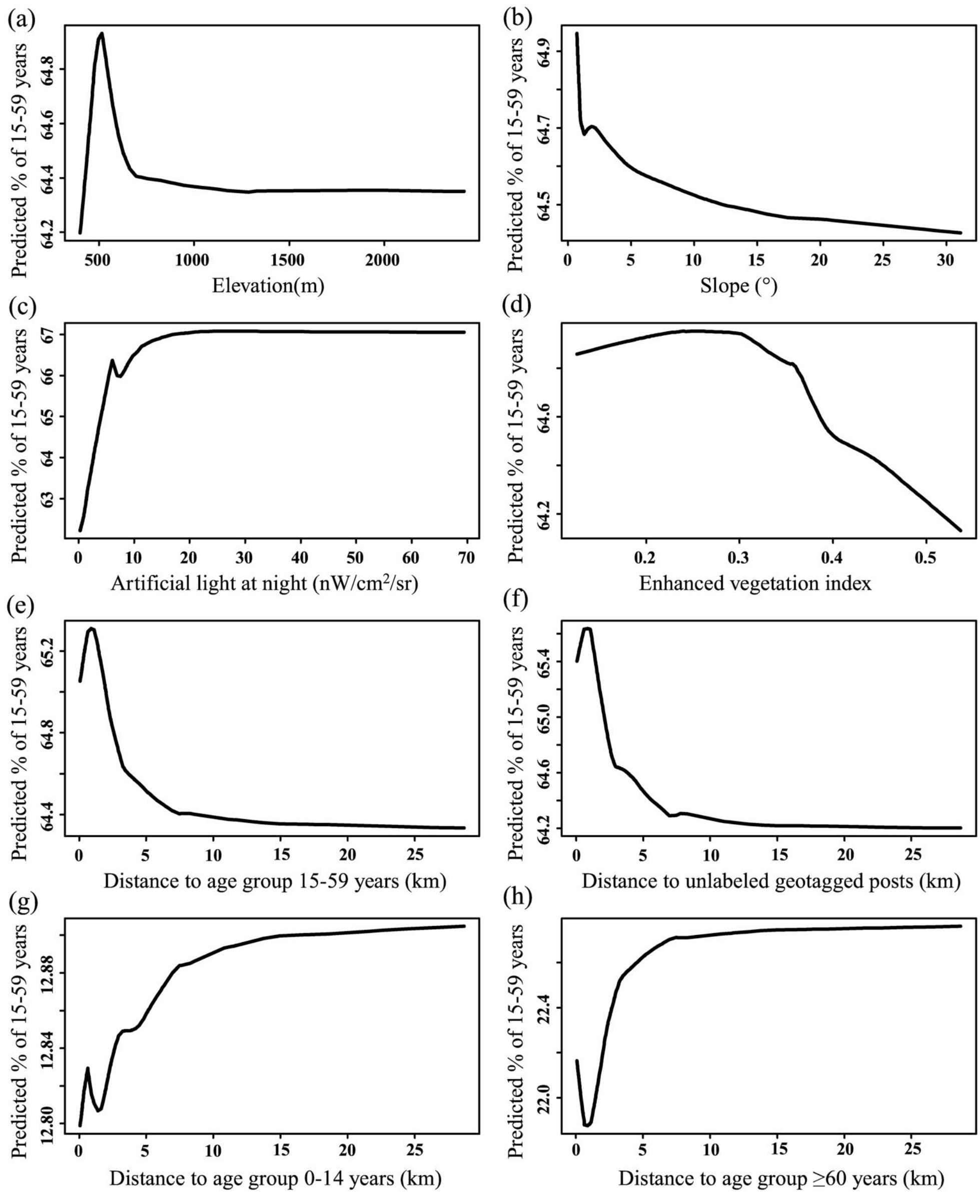

Figure 10. Associations between % of age group 15–59 years and key independent variables. 'Distance to *x*' denotes the distance to the nearest geotagged post labeled *x*.

group 0–14 years and the KDE- and IDW-based age-related deep features are shown in Figure S2.

The association between % of age group 15–59 years and elevation was found generally to be an inverse V-shaped curve (Figure 10). The % of age group 15–59 years first increased sharply to the top, then dropped about half and kept flattening as elevation increased. Slope was negatively associated with the % of age group 15–59 years. However, a positive association was observed between the % of age group 15–59 years and ALAN. The % of age group 15–59 years was found to increase slightly

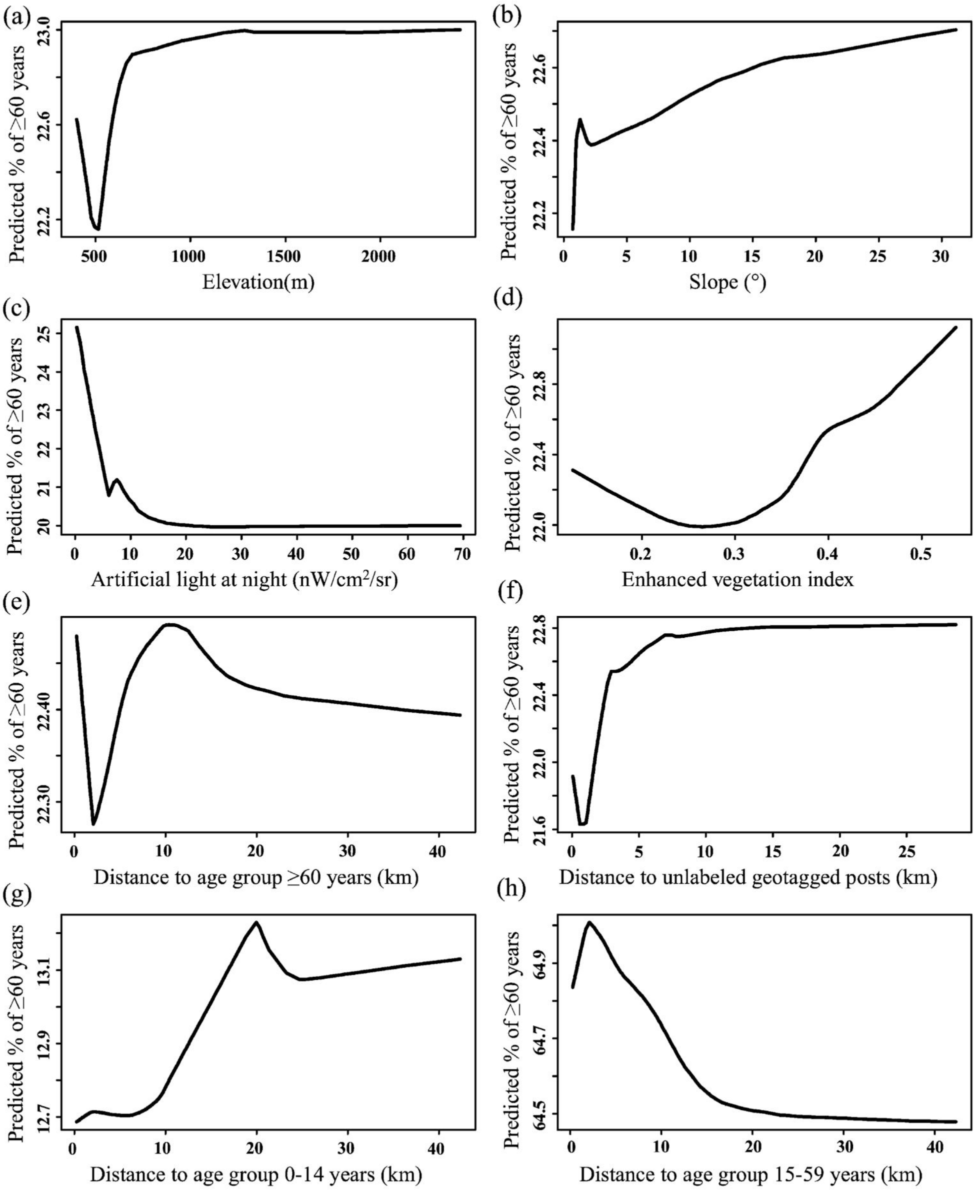

Figure 11. Associations between % of age group ≥60 years and key independent variables. 'Distance to *x*' denotes the distance to the nearest geotagged post labeled *x*.

before EVI rose to 0.3, and then to decrease rapidly. The % of age group 15–59 years was negatively associated with both the distance to the nearest geotagged post labeled 15–59 years and the distance to the nearest unlabeled geotagged post. The associations between the % of age group 15–59 years and the KDE- and IDW-based age-related deep features are shown in Figure S3.

The association between % of age group ≥60 years and elevation displayed a V-shaped curve (Figure 11). The % of age group ≥60 years first decreased and then increased rapidly around 500 m. After that, it slightly increased and became flat. Steeper slopes were generally associated with higher % of age group $\geq$60 years. Conversely, ALAN was negatively associated with the % of age group $\geq$60 years. As EVI increased, the % of age group $\geq$60 years declined slightly and then increased rapidly. A non-linear association between the distance to the nearest geotagged post labeled $\geq$60 years and % of age group $\geq$60 years was observed. The % of age group $\geq$60 years first decreased and then increased rapidly before 10 km, and then steadily decreased. The % of age group $\geq$60 years tends to be positively associated with the distance to the nearest unlabeled geotagged post. The associations of % of age group 15–59 years and % of age group $\geq$60 years with these key independent variables were almost inverse, as the Pearson correlation coefficient between them was found to be −0.95. The associations between the % of age group $\geq$60 years and the KDE- and IDW-based age-related deep features are shown in Figure S4.

5. Discussion

5.1. Summary of the key findings

This study developed a novel approach to estimate sex ratio, % of age groups 0–14, 15–59, and $\geq$60 years at a 100-m resolution, addressing the lack of fine-scale sex- and age-specific information in existing gridded population products. The fine-tuned BERT models and spatial modeling techniques were employed to extract and transform features from geotagged social media texts to gridded surfaces for predictive modeling. The interdependencies among demographic attributes were comprehensively captured using a multi-output random forest model with the Mahalanobis splitting rule. The demographic composition products generated by our modeling approach achieved higher accuracy than those produced by other approaches and the widely used WorldPop dataset. The variables that were important to the demographic compositions and their associations with demographic compositions were identified.

5.2. Comparisons of modeling approaches

Compared to classic machine learning models such as XGBoost, SVM, and FNN, which estimated demographic attributes separately without considering potential dependencies among them, our modeling approach achieved higher accuracy across most demographic attributes. This improvement can be attributed to the use of a tailored multi-output modeling approach that explicitly accounted for interdependencies among demographic attributes. Specifically, the proposed multi-output modeling approach employed different strategies for sex and age attributes based on their correlation structures. For the fully dependent sex attributes, their inherent dependence was addressed by consolidating them into a single variable – the sex ratio – which allowed for more efficient modeling without information loss. For the partially correlated age groups, a coherent multi-output random forest model was employed, which incorporated the covariance structure of age group percentages to capture the effect of independent variables on their joint variations, thereby improving both predictive stability and interpretability. Moreover, compared to MRF-ALL, which jointly modeled all demographic attributes at once, introducing complex constraints and irrelevant independent variables (e.g. sex-related variables may be weakly associated with age attributes), our approach strikes a better balance by explicitly considering key dependencies without incorporating excessive or irrelevant information.

Compared to the original BERT and LSTM models, features extracted by our fine-tuned BERT models resulted in more accurate estimates of gridded demographic compositions. The fine-tuned BERT models, trained on a small set of sex and age labels, updated the parameters of the original BERT to associate textual patterns with demographic characteristics. As such, the extracted deep features explicitly captured demographic information embedded in the text, unlike those from the original BERT, which merely reflected general language usage patterns. Although the LSTM models were also trained on labeled data, they relied on sequential processing and lacked the bidirectional attention mechanisms that enable models like BERT to capture comprehensive contextual relationships across the entire input. This limitation may restrict their capacity to recognize broader linguistic patterns relevant to demographic characteristics, making the extracted deep features less informative and leading to comparatively lower estimation accuracy.

Compared to the RS-only model, our approach using both remote sensing and social media variables achieved higher accuracy across all demographic attributes, demonstrating the added value of social media data for demographic composition estimation. The accuracy improvements varied across demographic attributes, with generally larger improvements observed for estimating the sex ratio than the % of different age groups. This difference may be related to the relatively lower performance of the fine-tuned BERT model in age-group classification, which could introduce additional uncertainty into the age-related information extracted from social media texts. Nevertheless, incorporating social media variables still consistently improved the estimation accuracy of all age groups. This suggests that social media variables remained informative for age-group estimation, providing supplementary information beyond that contained in remote sensing variables. Additionally, WorldPop was found to be less accurate than most of the evaluated approaches. This is likely because demographic compositions in WorldPop are directly derived from census data at relatively coarse administrative units (e.g. county level in China), rather than being estimated through fine-scale modeling with ancillary data, which limits its ability to capture intra-unit variability at the township level (i.e. the level used for relative accuracy assessment, and finer than the county level).

5.3. Interpretations of associations between independent variables and demographic compositions

Geotagged social media data played a pivotal role in estimating sex and age structures, with most of the top five important variables in estimating sex ratio, % of age groups 15–59, and $\geq$60 years being derived from Weibo data. The variables spatially processed by Euclidean distance and KDE were generally more important than those processed by IDW. As a spatial interpolation method, IDW may still require a certain spatial distribution of geotagged social media data for better performance (Shi *et al*. 2019). The distances to the nearest female- and male-labeled geotagged posts were the top two variables for estimating sex ratio. The positive association between the distance to the nearest female-labeled geotagged post and sex ratio suggests that increased distance from female users may indicate lower local presence of females, thereby contributing to higher sex ratios. It is also reasonable to observe the positive association between the distance to the nearest male-labeled geotagged post and sex ratio. In China, higher sex ratios have been reported in rural areas as compared to urban areas, and such rural areas typically exhibit greater distances to male-labeled posts due to the sparser spatial distribution of geotagged posts (Zhou *et al*. 2011, Jiang *et al*. 2017, Wang *et al*. 2021). Social media data generally contributed less to estimating % of age groups than to estimating sex ratio. The negative association between the distance to the nearest geotagged post labeled 15–59 years and the corresponding population percentage suggests that locations farther from users aged 15–59 years tend to have a smaller share of population within that age group. However, population bias of the social media users could lead to incomplete representation of children and the elderly (Montasser and Kifer 2017), leading to difficulties in explaining the observed positive association between the distance to the nearest geotagged post labeled 0–14 years and % of age group 0–14 years, and the non-linear relationship between the distance to the nearest geotagged post labeled $\geq$60 years and % of age group $\geq$60 years. In addition, the smaller number of age labels available for fine-tuning the BERT models, compared with sex labels, may limit the extraction of age-related information from social media data, leading to a lower contribution of social media variables to age-group estimation.

Remote sensing data could supplement social media data in the estimation of age structures. ALAN was the most important variable in estimating the % of age groups 15–59 and $\geq$60 years, and elevation was the second most influential variable in estimating the % of age group 0–14 years. Higher ALAN was associated with larger % of age groups 0–14 and 15–59 years, and a smaller % of age group $\geq$60 years. This trend was consistent with previous studies in China, where the elderly were more concentrated in economically underdeveloped areas (Cheng *et al*. 2019, Man *et al*. 2021). The similar reason could also explain the associations of EVI and slope with the proportions of working-age and older populations, as urban areas with stronger economies tend to exhibit lower average EVI and slope values compared to rural areas. The highest % of age group 15–59 years and the lowest % of age group $\geq$60 years were observed at elevations of around 500 m, which can be attributed to the fact that the economically developed urban areas of Chengdu City are located at this elevation, whereas high-altitude rural areas were associated with lower proportions of working-age people and higher proportions of older populations (Qin 2015). A similar urban–rural gradient may also help interpret the nonlinear associations between environmental variables and sex ratio, as the turning points in the associations – observed at around 750 m in elevation, 20 nW/cm^2/sr in ALAN, and 0.35 in EVI – correspond to urban-rural transition zones in Chengdu City. In these urban-rural transition zones, the coexistence of urban and rural characteristics may lead to larger and more complex variations in sex ratio than in predominantly urban or rural areas. As such, the environmental factors derived from remote sensing data could serve as surrogates for factors that influence demographic compositions (Alegana *et al.* 2015).

5.4. Strengths and limitations

This study presents three major advantages. First, a novel approach was developed to, for the first time, estimate complete sex and age structures at high spatial resolution, which capture spatial heterogeneities within even the smallest administrative units and can be readily integrated with other geospatial data for broader applications. Second, this study is the first to employ social media data for estimating demographic compositions at the grid level. The demographics-related variables derived from social media data provided direct measurements of individual demographic attributes, offering explicit and subjectively measured information on demographic compositions across geography. These variables supplement environmental variables derived from remote sensing data, which are objectively measured physical proxies that correlate with, but do not directly represent, demographic compositions. The integration of these two data sources enables more comprehensive and nuanced estimation of demographic compositions. Furthermore, their broad availability makes our approach potentially applicable even in resource-poor settings. Third, our multi-output modeling strategy effectively considered dependencies among multiple demographic attributes, leading to efficient and accurate estimates. Compared to previous approaches that first estimate the population densities of different demographic groups and then derive their proportions, our approach directly estimates demographic compositions, which reduces uncertainty propagation and more explicitly captures the dependencies among demographic attributes (Li *et al.* 2016, Zhao *et al.* 2021b, Ju *et al.* 2025). Although deep learning methods inherently account for dependencies among multiple outputs, they typically require large amounts of training data (Jeong and Lee 2026), whereas the random forest models used in this study are more data-efficient and can accommodate varying sample sizes, making them suitable for estimating demographic compositions at both large and small scales.

There are also limitations in this study. First, although the fine-tuned BERT models extracted important and influential features from the social media texts for demographic compositions modeling, linking these deep features to clear and definitive real-world interpretations remains challenging due to the limited interpretability of deep learning models (Loyola-González 2019). Although exploratory interpretations were performed to provide indicative semantic insights, these interpretations are based on associative patterns and do not necessarily reflect the actual semantic meanings of the deep features. Future studies could further examine potential drivers of demographic compositions to enhance the interpretability of text-derived deep features. Second, social media data are subject to population bias, as younger individuals tend to be more active on social media, and selective geotagging can occur. Although this bias remains a concern, our modeling approach may partially mitigate its impact through multiple strategies: by using IDW, KDE, and Euclidean distance to spatially process the point-based social media data, the influence of locally over-represented biased samples could be smoothed by nearby samples; environmental factors from remote sensing data were employed, which are objectively measured and provide unbiased predictors that could reduce reliance on the potentially biased social media variables; and by modeling multiple demographic attributes jointly, our approach leveraged dependencies among outputs, which may reduce the influence of biased variables on any specific demographic attributes. While these strategies may partially mitigate the effects of population bias, future studies should further assess the representativeness of social media data and explore corrective strategies, such as stratified sampling, to further mitigate this issue. Third, the relatively small number of geotagged social media data may constrain the accuracy of demographic composition estimates. However, social media data directly measure individual demographic characteristics in given areas, offering valuable demographic indicators that explicitly link demographic compositions to geographic locations, which, to some extent, remain informative even when sparsely distributed. Additionally, despite not covering every 100-m grid, the geotagged social media data used in this study provided substantial spatial coverage of densely populated urban areas, providing valuable information where demographic compositions are most needed. Leveraging multiple spatial modeling techniques, including KDE, IDW, and Euclidean distance, these social media data can support reliable spatial extrapolation, enabling the derived spatial variables to effectively capture heterogeneities even in areas not directly covered by the data. Future studies could incorporate larger volumes of social media data to further enhance estimation accuracy and support the estimation of more detailed demographic compositions, such as finer age group divisions. Fourth, the associations between demographic compositions and ancillary data at the township level were used to generate grid-level estimations. This approach could lead to over- or under-estimation of demographic compositions, as the averaged values in the coarser unit could not fully describe the variation of variables within finer units (Sinha *et al*. 2019, Mei *et al*. 2022). Using fine-scale micro census data as reference of demographic compositions could be an option to reduce such uncertainties (Boo *et al*. 2022).

5.5. Usefulness of gridded demographic composition data

Sex and age are among the most fundamental demographic attributes, as they reveal the relative proportions of young and old as well as the balance of males and females, which underpin various essential social behaviors such as schooling, employment, marriage, and support for the elderly (Kahn 2007). The gridded datasets generated by our modeling approach offer fine-resolution, standardized, and georeferenced cells of sex and age structures, which are crucial for planning and resource allocation across multiple sectors. In public health, detailed sex and age structures enable more precise healthcare planning (Jia *et al*. 2019). For instance, maternal and child healthcare professionals can be allocated to hospitals and clinics with higher demand possibly due to larger populations of children and women of childbearing age (Tatem *et al*. 2014). Schools and nursing homes can be strategically located in areas with higher concentration of children and the elderly to enhance social services (You *et al*. 2017).

Additionally, entertaining goods can be directed to retailers in regions where young adults comprise a larger share of the population, improving marketing effectiveness (Barr 2004).

6. Conclusions

This study introduced a novel approach for estimating complete sex and age structures on 100-m grid cells, utilizing readily accessible remote sensing and social media data. The state-of-the-art BERT model was fine-tuned to extract age- and sex-related features from social media texts. Euclidean distance, KDE, and IDW then estimated these features across grids for fine-scale modeling. The correlation among multiple demographic attributes was considered by the multi-output random forest using the Mahalanobis splitting rule. The resulting products from our approach outperformed the widely used WorldPop age and sex structure dataset. Both remote sensing and social media ancillary data were important during modeling, with social media data more useful in estimating sex structures. Our modeling approach not only enables accurate estimation of high-resolution demographic compositions but also holds promise for generating other spatial products that rely on multimodal data and require multiple outputs, benefiting various fields dependent on these essential datasets. Ultimately, this work contributes to a deeper understanding of population structure and dynamics and supports more informed decision-making in areas such as urban planning, public health, and resource allocation.

Acknowledgments

We sincerely thank the editors and the anonymous reviewers for their valuable comments and suggestions.

Author contributions

CRediT: **Ge Qiu**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Yuchen Li**: Methodology, Validation, Writing – original draft, Writing – review & editing; **Shaoqing Dai**: Investigation, Validation, Visualization, Writing – review & editing; **Kun Qin**: Data curation, Software, Visualization, Writing – review & editing; **Zhanpeng Wang**: Data curation, Investigation, Visualization, Writing – review & editing; **Yaolin Liu**: Conceptualization, Project administration, Resources, Supervision, Validation, Writing – review & editing; **Jiping Liu**: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing; **Peng Jia**: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

AI disclosure statement

The authors used ChatGPT (version 5.1) solely for the purpose of language editing and polishing the manuscript.

Funding

This work was supported by the National Key R&D Program of China (2023YFC3604704 and 2022YFC3005705), National Natural Science Foundation of China (42271433), Fundamental Research Funds for the Central Universities (413000159 and 2042024kf1024), and the International Institute of Spatial Lifecourse Health (ISLE).

Notes on contributors

Ge Qiu received his Ph.D. in Cartography and Geographic Information Systems from the School of Resource and Environmental Sciences, Wuhan University. His research interests are multi-source geospatial data fusion, spatial modeling and mapping, GeoAI, and health geography. His primary contributions to this study included conceptualization, methodology, investigation, formal analysis, software, validation, visualization, data curation, writing – original draft, and writing – review and editing.

Yuchen Li is a Lecturer at the School of Geography, University of Leeds. His research interests are spatial analysis and modeling, spatial data mining, spatial statistics, transportation simulation, and public health modeling. His primary contributions to this study included methodology, validation, writing – original draft, and writing – review and editing.

Shaoqing Dai is a postdoctoral researcher at the School of Resource and Environmental Sciences, Wuhan University. His research interests are health geography, spatial statistics, GeoAI, and urban visual intelligence. His primary contributions to this study included investigation, validation, visualization, and writing – review and editing.

Kun Qin is a Ph.D. candidate at the School of Resource and Environmental Sciences, Wuhan University. His research interests include spatial modeling and mapping, spatiotemporal analysis, and health geography. His primary contributions to this study included software, visualization, data curation, and writing – review and editing.

Zhanpeng Wang received his Ph.D. in Cartography and Geographic Information Systems from the School of Resource and Environmental Sciences, Wuhan University. His research interests are remote sensing, GeoAI, and health geography. His primary contributions to this study included investigation, visualization, data curation, and writing – review and editing.

Yaolin Liu is a Professor at the School of Resource and Environmental Sciences, Wuhan University, and currently serves as the Chancellor of Duke Kunshan University. His research focuses on the application of geospatial technologies in land resource surveying, evaluation, planning, and management. His primary contributions to this study included conceptualization, validation, resources, supervision, project administration, and writing – review and editing.

Jiping Liu is a Research Fellow and Deputy Director General at the Chinese Academy of Surveying and Mapping. His research interests are spatiotemporal big data analysis and mining, governmental geospatial information services, emergency geospatial information services, and 3D real-world modeling. His primary contributions to this study included conceptualization, investigation, validation, resources, supervision, funding acquisition, project administration, and writing – review and editing.

Peng Jia is the founding director of the International Institute of Spatial Lifecourse Health (ISLE) and the Deputy Director of Medical Remote Sensing Information Research Institute (MRSIRI) at

Renmin Hospital of Wuhan University. His research interests include health geography, spatial epidemiology, environmental health, and spatial science and technology. His primary contributions to this study included conceptualization, validation, resources, supervision, funding acquisition, project administration, and writing – review and editing.

Data and codes availability statement

The data and codes that support the findings of this study are available in: https://doi.org/10.6084/m9.figshare.32826434

References

Abdul-Mageed, M., *et al.*, 2019. *Sentence-level BERT and multi-task learning of age and gender in social media* [online]. Available from: https://ui.adsabs.harvard.edu/abs/2019arXiv191100637A [Accessed 1 November 2019].

Abdullah, A.M., *et al.*, 2022. Age- and sex-specific association between vegetation cover and mental health disorders: Bayesian spatial study. *JMIR Public Health and Surveillance*, 8 (7), e34782.

Alakus, C., Larocque, D., and Labbe, A., 2023. Covariance regression with random forests. *BMC Bioinformatics*, 24 (1), 258.

Alegana, V.A., *et al.*, 2015. Fine resolution mapping of population age-structures for health and development applications. *Journal of the Royal Society Interface*, 12 (105), 20150073.

Asal, F.F., 2012. Visual and statistical analysis of digital elevation models generated using Idw interpolator with varying powers. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-2, 57–62.

Bakillah, M., *et al.*, 2014. Fine-resolution population mapping using OpenStreetMap points-of-interest. *International Journal of Geographical Information Science*, 28 (9), 1940–1963.

Balk, D., and Yetman, G., 2004. *The global distribution of population: evaluating the gains in resolution refinement*. New York: Center for International Earth Science Information Network, Columbia University.

Balk, D.L., *et al.*, 2006. Determining global population distribution: methods, applications and data. *In:* S. I. Hay, A. Graham, and D. J. Rogers, eds. Advances in *parasitology*, Vol 62: *global mapping of infectious diseases: methods, examples and emerging applications*. Amsterdam, The Netherlands: Elsevier, 119–156.

Barr, T.F., 2004. Demographic targeting: the essential role of population groups in retail marketing. *Journal of Consumer Marketing*, 21, 67–69.

Boo, G., *et al.*, 2022. High-resolution population estimation using household survey data and building footprints. *Nature Communications*, 13 (1), 1330.

Breiman, L., 2001. Random forests. *Machine Learning*, 45 (1), 5–32.

Calderón-Argelich, A., *et al.*, 2023. Greening plans as (re)presentation of the city: toward an inclusive and gender-sensitive approach to urban greenspaces. *Urban Forestry & Urban Greening*, 86, 127984.

Castronuovo, L., *et al.*, 2021. Food marketing and gender among children and adolescents: a scoping review. *Nutrition Journal*, 20 (1), 52.

Chen, T., *et al.*, 2022. xgboost: Extreme Gradient Boosting. R package version 1.6.0.1. https://CRAN.R-project.org/package=xgboost.

Cheng, Y., *et al.*, 2019. Understanding the spatial disparities and vulnerability of population aging in China. *Asia & the Pacific Policy Studies*, 6 (1), 73–89.

Cui, L., *et al.*, 2018. Prediction of the healthcare resource utilization using multi-output regression models. *IISE Transactions on Healthcare Systems Engineering*, 8 (4), 291–302.

Devlin, J., *et al.*, 2019. BERT: pre-training of deep bidirectional transformers for language understanding. *In: Conference of the North-American-chapter of the association-for-computational-*

linguistics - human language technologies (NAACL-HLT), Jun 02-07 2019. Minneapolis, MN: Association for Computational Linguistics, 4171–4186.

Didan, K., 2015. MOD13Q1 MODIS/Terra vegetation indices 16-day L3 global 250m SIN grid V006. *NASA Eosdis Land Processes Daac*, 10 (730), 415.

Dobson, J.E., *et al.*, 2000. LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, 66 (7), 849–857.

Drefahl, S., *et al.*, 2020. A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. *Nature Communications*, 11 (1), 5097.

Elvidge, C.D., *et al.*, 2021. Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, 13 (5), 922.

Ethayarajh, K. 2019. How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. *In:* K. Inui, *et al.*, eds., *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, November 2019 Hong Kong, China, 55–65.

Evgeniou, T., and Pontil, M., 2004. Regularized multi–task learning. *In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining*. Seattle, WA, USA: Association for Computing Machinery, 109–117.

Farr, T.G., and Kobrick, M., 2000. Shuttle radar topography mission produces a wealth of data. *Eos, Transactions American Geophysical Union*, 81 (48), 583–585.

Fecht, D., *et al.*, 2020. Advances in mapping population and demographic characteristics at small-area levels. *International Journal of Epidemiology*, 49 (Suppl 1), I15–I25.

Feng, Q.S., *et al.*, 2019. Age of retirement and human capital in an aging China, 2015–2050. *European Journal of Population = Revue Europeenne de Demographie*, 35 (1), 29–62.

Freire, S., *et al.*, 2016. Development of new open and free multi-temporal global population grids at 250 m resolution. *In: Geospatial data in a changing world*. Helsinki, Finland: Association of Geographic Information Laboratories in Europe (AGILE), 119–128.

Gao, Q., *et al.*, 2012. A comparative study of users' microblogging behavior on Sina Weibo and Twitter. *In:* J. Masthoff, *et al.*, eds., *User modeling, adaptation, and personalization*. Berlin, Heidelberg: Springer, 88–101.

Gatrell, A.C., *et al.*, 1996. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, 21 (1), 256–274.

Gebru, T., *et al.*, 2017. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 114 (50), 13108–13113.

Guimaraes, R.G., *et al.*, 2017. Age groups classification in social network using deep learning. *IEEE Access*, 5, 10805–10816.

Hochreiter, S., and Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, 9 (8), 1735–1780.

Huang, Y.Y., *et al.*, 2022. Exploring the relationship between the spatial distribution of different age populations and points of interest (POI) in China. *ISPRS International Journal of Geo-Information*, 11 (4), 215.

Ishwaran, H., *et al.*, 2021. randomForestSRC: multivariate splitting rule vignette.

Jeong, B., and Lee, B.K., 2026. Popnet: computer vision based deep learning model for forecasting gridded population. *International Journal of Geographical Information Science*, 40 (1), 217–236.

Jia, P., Qiu, Y., and Gaughan, A.E., 2014. A fine-scale spatial population distribution on the High-resolution Gridded Population Surface and application in Alachua County, Florida. *Applied Geography*, 50 (2), 99–107.

Jia, P., Shi, X., and Xierali, I.M., 2019. Teaming up census and patient data to delineate fine-scale hospital service areas and identify geographic disparities in hospital accessibility. *Environmental Monitoring and Assessment*, 191 (S2), 1–14.

Jiang, Q.B., *et al.*, 2017. Changes in sex ratio at birth in china: a decomposition by birth order. *Journal of Biosocial Science*, 49 (6), 826–841.

Jiang, W., *et al.*, 2021. Using geotagged social media data to explore sentiment changes in tourist flow: a spatiotemporal analytical framework. *ISPRS International Journal of Geo-Information*, 10 (3), 135.

Ju, Y., *et al.*, 2025. 100-m resolution age-stratified population estimation from the 2020 China census by township (ASPECT). *Scientific Data*, 12 (1), 1058.

Kahn, J.R., 2007. *Demographic techniques: population pyramids and age/sex structure*. Malden, MA: The Blackwell Encyclopedia of Sociology.

Kontur, D., 2024. HDX: Kontur population - global population density for 400 m H3 hexagons.

Li, S., *et al.*, 2016. Population and age structure in Hungary: a residential preference and age dependency approach to disaggregate census data. *Journal of Maps*, 12 (sup1), 560–569.

Liaw, A., and Wiener, M., 2002. Classification and regression by randomForest. *R News*, 2 (3), 18–22.

Loyola-González, O., 2019. Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view. *IEEE* Access, 7, 154096–154113.

Lu, G.Y., and Wong, D.W., 2008. An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences*, 34 (9), 1044–1055.

Man, W., Wang, S.B., and Yang, H., 2021. Exploring the spatial-temporal distribution and evolution of population aging and social-economic indicators in China. *BMC Public Health*, 21 (1), 966.

Mei, Y.A., *et al.*, 2022. Population spatialization with pixel-level attribute grading by considering scale mismatch issue in regression modeling. *Geo-Spatial Information Science*, 25 (3), 365–382.

Merchant, A., *et al.*, 2020. What happens to BERT embeddings during fine-tuning? *In:* A. Alishahi, *et al.*, eds., *Proceedings of the third BlackboxNLP workshop on analyzing and interpreting neural networks for NLP*, November 2020, 33–44.

Meyer, D., *et al.*, 2023. e1071: misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien. R package version 1.7-13. https://CRAN.R-project.org/package=e1071.

Miao, R., Wang, Y., and Li, S., 2021. Analyzing urban spatial patterns and functional zones using Sina Weibo POI data: a case study of Beijing. *Sustainability*, 13 (2), 647.

Mislove, A., *et al.*, 2021. Understanding the demographics of Twitter users. *Proceedings of the International AAAI Conference on Web and Social Media*, 5 (1), 554–557.

Montasser, O., and Kifer, D., 2017. Predicting demographics of high-resolution geographies with geotagged tweets. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31 (1), 1460–1466.

National Bureau of Statistics. 2021. *Communiqué of the seventh national population census (No.* 5) [online]. National Bureau of Statistics, PRC. Available from: https://www.stats.gov.cn/english/PressRelease/202105/t20210510_1817190.html.

Nguyen, D., *et al.*, 2021. "How old do you think I am?" a study of language and age in Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 7 (1), 439–448.

Nia, Z.M., *et al.*, 2023. Twitter-based gender recognition using transformers. *Mathematical Biosciences and Engineering: MBE*, 20 (9), 15962–15981.

Nilsen, K., *et al.*, 2021. A review of geospatial methods for population estimation and their use in constructing reproductive, maternal, newborn, child and adolescent health service indicators. *BMC Health Services Research*, 21 (SUPPL 1), 370.

O'connor, K., *et al.*, 2024. Methods and annotated data sets used to predict the gender and age of Twitter users: scoping review. *Journal of Medical Internet Research*, 26, e47923.

Pezzulo, C., *et al.*, 2017. Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. *Scientific Data*, 4 (1), 170089.

Qin, B., 2015. City profile: Chengdu. *Cities*, 43, 18–27.

Qin, K., *et al.*, 2024. Spatiotemporal patterns of air pollutants over the epidemic course: a national study in China. *Remote Sensing*, 16 (7), 1298.

Qiu, G., *et al.*, 2020. Local population mapping using a random forest model based on remote and social sensing data: a case study in Zhengzhou, China. *Remote Sensing*, 12 (10), 1618.

Qiu, G., *et al.*, 2025. High-resolution population density mapping in urban areas using a contextualized geographically weighted neural network (CGWNN) model. *Applied Geography*, 182, 103708.
Rahman, R., Otridge, J., and Pal, R., 2017. IntegratedMRF: random forest-based framework for integrating prediction from different data types. *Bioinformatics (Oxford, England)*, 33 (9), 1407–1410.
Scarpone, C., *et al.*, 2020. A multimethod approach for county-scale geospatial analysis of emerging infectious diseases: a cross-sectional case study of COVID-19 incidence in Germany. *International Journal of Health Geographics*, 19 (1), 32.
Segal, M., and Xiao, Y., 2011. Multivariate random forests. *WIREs Data Mining and Knowledge Discovery*, 1 (1), 80–87.
Shi, X., *et al.*, 2019. Estimation of environmental exposure: interpolation, kernel density estimation or snapshotting. *Annals of GIS*, 25 (1), 1–8.
Sinha, P., *et al.*, 2019. Assessing the spatial sensitivity of a random forest model: application in gridded population modeling. *Computers Environment and Urban Systems*, 75, 132–145.
Steele, J.E., *et al.*, 2017. Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface*, 14 (127), 20160690.
Strobl, C., *et al.*, 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8 (1), 25.
Suel, E., *et al.*, 2019. Measuring social, environmental and health inequalities using deep learning and street imagery. *Scientific Reports*, 9 (1), 6229.
Sun, C., *et al.*, 2019. How to fine-tune bert for text classification? *In: Chinese computational linguistics: 18th China national conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18*, 194–206.
Szabo, S., *et al.*, 2020. Health workforce demography: a framework to improve understanding of the health workforce and support achievement of the Sustainable Development Goals. *Human Resources for Health*, 18 (1), 7.
Tang, F., and Ishwaran, H., 2017. Random forest missing data algorithms. *Statistical Analysis and Data Mining*, 10 (6), 363–377.
Tatem, A.J., *et al.*, 2014. Mapping for maternal and newborn health: the distributions of women of childbearing age, pregnancies and births. *International Journal of Health Geographics*, 13 (1), 2.
Tatem, A.J., *et al.*, 2013. Quantifying the effects of using detailed spatial demographic data on health metrics: a systematic analysis for the AfriPop, AsiaPop, and AmeriPop projects. *The Lancet*, 381, S142.
Tiecke, T.G., *et al.*, 2017. *Mapping the world population one building at a time* [online]. Available from: https://ui.adsabs.harvard.edu/abs/2017arXiv171205839T [Accessed 1 December 2017].
Tobler, W.R., 1970. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46 (sup1), 234–240.
Tukey, J.W., 1977. *Exploratory data analysis*. Reading, MA: Addison-Wesley Publishing Company.
United Nations. 2015. *Transforming our world: the 2030 Agenda for Sustainable Development: resolution/adopted by the General Assembly*.
Vaswani, A., *et al.*, 2017. Attention is all you need. *In: 31st annual conference on neural information processing systems (NIPS)*, Dec 04-09 2017. Long Beach, CA: Advances in Neural Information Processing Systems.
Venables, W.N., and Ripley, B.D., 2002. *Modern applied statistics with S*. New York: Springer.
Wang, J., *et al.*, 2021. Mapping the exposure and sensitivity to heat wave events in China's megacities. *The Science of the Total Environment*, 755 (Pt 1), 142734.
Wu, C., *et al.*, 2018. Check-in behaviour and spatio-temporal vibrancy: an exploratory analysis in Shenzhen, China. *Cities*, 77, 104–116.
Xu, Y.M., *et al.*, 2021. Comparative assessment of gridded population data sets for complex topography: a study of Southwest China. *Population and Environment*, 42 (3), 360–378.
Yang, Q.Q., *et al.*, 2023. A synchronized estimation of hourly surface concentrations of six criteria air pollutants with GEMS data. *NPJ Climate and Atmospheric Science*, 6 (1), 94.
Yang, X.C., *et al.*, 2019. Population mapping with multisensor remote sensing images and point-of-interest data. *Remote Sensing*, 11 (5), 574.

Ye, T.T., *et al.*, 2019. Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. *The Science of the Total Environment*, 658, 936–946.
You, N.L., Shen, Z.J., and Nishino, T., 2017. Assessing the allocation of special elderly nursing homes in Tokyo, Japan. *International Journal of Environmental Research and Public Health*, 14 (10), 1102.
Zhang, J.Y., and Zhao, X.S., 2024. Using POI and multisource satellite datasets for mainland China's population spatialization and spatiotemporal changes based on regional heterogeneity. *The Science of the Total Environment*, 912, 169499.
Zhao, X., *et al.*, 2021a. Mapping population distribution based on XGBoost using multisource data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 11567–11580.
Zhao, Y.C., *et al.*, 2021b. Intraday variation mapping of population age structure via urban-functional-region-based scaling. *Remote Sensing*, 13 (4), 805.
Zhou, X.D., *et al.*, 2011. The very high sex ratio in rural China: impact on the psychosocial well-being of unmarried men. *Social Science & Medicine (1982)*, 73 (9), 1422–1427.
Zhu, Z.X., and Mao, K.Z., 2023. Knowledge-based BERT word embedding fine-tuning for emotion recognition. *Neurocomputing*, 552, 126488.
Zhuang, H.M., *et al.*, 2021. Mapping multi-temporal population distribution in china from 1985 to 2010 using landsat images via deep learning. *Remote Sensing*, 13 (17), 3533.