Using Satellite Images and Deep Learning to Identify Associations Between County-Level Mortality and Neighborhood Features

Joshua J. Levy, B.A.[1,2,3,*], Rebecca M. Lebeaux, B.S.[1,2], Anne G. Hoen, Ph.D.[1,2,4], Brock C. Christensen, Ph.D.[1], Louis J. Vaickus, MD Ph.D.[3], Todd A. MacKenzie, Ph.D.[1,4,5]

1. Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA
2. Program in Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA
3. Department of Pathology, Dartmouth Hitchcock Medical Center
4. Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA
5. The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA

*To whom correspondence should be addressed: joshua.j.levy.gr@dartmouth.edu

Abstract

What is the relationship between mortality and satellite images as elucidated through the use of Convolutional Neural Networks?

Background: Following a century of increase, life expectancy in the United States has stagnated and begun to decline in recent decades. Using satellite images and street view images, prior work has demonstrated associations of the built environment with income, education, access to care and health factors such as obesity. However, assessment of learned image feature relationships with variation in crude mortality rate across the United States has been lacking. We sought to investigate if county-level mortality rates in the U.S. could be predicted from satellite images.

Methods: Satellite images were extracted with the Google Static Maps application programming interface for 430 counties representing approximately 68.9% of the US population. A convolutional neural network was trained using crude mortality rates for each county in 2015 to predict mortality. Learned image features were interpreted using Shapley Additive Feature Explanations, clustered, and compared to mortality and its associated covariate predictors.

Results: Predicted mortality from satellite images in a held-out test set of counties was strongly correlated to the true crude mortality rate (Pearson r=0.72). Learned image features were clustered, and we identified 10 clusters that were associated with education, income, geographical region, race and age. Direct prediction of mortality using a deep learning model across a cross-section of 430 U.S. counties identified key features in the environment (e.g. sidewalks, driveways and hiking trails) associated with lower mortality.

Conclusions: The application of deep learning techniques to remotely-sensed features of the built environment can serve as a useful predictor of mortality in the United States. Although we identified features that were largely associated with demographic information, future modeling approaches that directly identify image features associated with health-related outcomes have the potential to inform targeted public health interventions.

Introduction

Life expectancy in the United States has increased dramatically over the past century from 48 years in 1900 to 80 years in 2019. However, the United States has experienced a drop in longevity over the past decade and now ranks 43rd in the world [1–5]. Within the United States, crude mortality rates vary by more than 40%. Factors observed to be associated with mortality rates within the United States include disparities in socioeconomic status[6, 7] and health insurance coverage,[8] as well as obesity, smoking[9, 10] and drug use/opioid abuse[11].

Prior studies have attempted to infer characteristics of the underlying communities by characterizing land use using satellite imagery[12–14]. Recent research has leveraged deep learning approaches to link the built environment to obesity[15, 16], socioeconomic status[17], poverty[18, 19], and other demographic factors [20, 21] by integrating lower level image features into higher order abstractions to make predictions [22, 23]. In addition, the potential for deep learning-based image analysis to characterize a broad range of health exposures was recently reviewed [24]. However, there is limited work that examines the relationship of mortality with the built environment in the United States at a large scale. The identification of information in satellite images associated with mortality predictions could potentially unlock previously unknown, yet related geographic and structural community characteristics that may be used to predict future mortality rates when demographic information is unavailable or possibly inform optimal county/state public health intervention strategies.

We hypothesized that county-scale level mortality rates could be predicted from satellite images. The goals of this study were to determine whether the analysis of satellite images by Convolutional Neural Networks (CNN) could be used to predict county-level mortality rates and uncover salient satellite image features associated with mortality. We also sought to determine if image features are related with county-level measures of income, education, age, sex, race and ethnicity. The main objective of

this study was to highlight a proof-of-concept deep learning application that presents strong baseline performance, enabling future work that can evaluate the potential for applying learnt features to the design of public health interventions in the built environment.

Methods

Data Collection

To train our deep learning model, we used publicly available county-level mortality data. We used the CDC Wonder database to collect all available death data from 2015 for residents of the 50 United States and the District of Columbia, and matched death data with 2015 county Census population data from the Bureau of Economic Analysis, USDA ERS databases[25–27] , and the Surveillance, Epidemiology, and Ends Results (SEER) Program[28]. Crude mortality rates were calculated as the number of deaths in the county divided by the population in the county. Additional county level covariate information from 2015 was collected on age, sex, Hispanic status, race, income, education, and region (Supplementary Table 1), though this covariate information was not used in our computer vision model.

We selected 430 counties from among 3,142 total United States counties in 2015 for inclusion in our study. This subset of representative counties was selected to decrease the computational burden and data storage resources. To reduce the variance of the mortality estimate and to limit potential bias from locale (e.g. rural, urban), we selected the top 1000 most populous counties (restricting to more urban environments) and split these counties into 13 bins containing similar numbers of counties rank ordered by mortality. We selected up to forty of the most populous counties in each bin. These selection criteria were developed to capture a greater proportion of total deaths and represent wider variation in mortality rates. 279 counties were placed into the training set (65%), 65 in the validation set (15%), and 86 in the test set (20%). The training set was used to update the parameters of the trained model while the validation set was used to

limit the model from overfitting to the training data (Figure 1a-b). The held-out test set represents an application of the modeling approach to unseen counties that had no role in the training of the model.

Satellite imagery data was collected using the Google Static Maps application programming interface (API) to build our deep learning pipeline, similar to Maharana and Nsoesie[15]. First, four schools from each county were randomly selected to serve as points of interest (POI) to sample nearby images. We used schools because they are typically placed in densely populated regions of the county and to reduce the computing space usage and required resources at our local computing cluster and acknowledge that the selection of schools as a POI presents a limited view of the entire county. We obtained geographic coordinates for the schools in our study from the National Center for Education Statistics[29], and divided the 1 square mile area surrounding each school into an evenly spaced 7 images by 7 images grid (extracting 196 total images per county) (Figure 1c). Preliminary tests from our modeling approach indicated that mortality prediction performance increases with the number of schools and area around the school sampled. However, performance saturates as the number of images in each county approaches 196, thereby warranting selection of images contained within a one-mile square area around schools to maximize the potential utility of assessing schools as a POI and limit the amount of noise created beyond surveying schools. The zoom level for each image was set to 17, and image dimensions were set to 400 pixels by 400 pixels to provide enough detail to make out street patterns and cover the space between the selected images of the grid. 84,280 satellite images were downloaded using the Google Static Maps API. Images were downloaded between July and September 2019, representing a collection of images acquired April 2018 to December 2018.

This study was exempt from institutional review board approval as we accessed previously collected data and could not identify individuals. Our study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross-sectional studies where appropriate[30].

Deep Learning Model Training

Deep neural networks have been used in a wide range of health-related applications[31, 32]. Convolutional Neural Networks (CNN) slide filters across images to pick up on low level features such as edges or curves and then expand the visual field to pick up higher-order constructs (Figure 1c) [33]. The particular variant of CNN that was used to predict mortality on the satellite images was ResNet-50 [34], pretrained on the ImageNet database to recognize over 1000 different objects [35], some of which may have features corresponding with our satellite image set. We performed transfer learning, which applies knowledge gained from ImageNet to initialize the parameters of our model. These model parameters were updated to minimize the divergence between the true and predicted mortality rates using the negative Poisson log likelihood; training the model with this objective allowed us to use each image to directly predict the mortality of its respective county. Training images were randomly cropped, rotated, resized, and flipped to improve the generalizability of the approach. Predictions averaged across the images for each county comprised the final prediction. After training our model on images from 279 counties, we evaluated the model on images from the remaining test counties and averaged the predicted mortality rates across each county to derive county mortality estimates. The validation set was used to terminate the learning process at 5 training iterations and identify the ideal learning rate for the model, 1e-4. The deep learning models developed using Python 3.6, utilized the PyTorch 1.3.0 framework and were trained using K80 Graphics Processing Units (GPUs). A sensitivity analysis was utilized to decide the ideal number of schools and sampling area around the schools. We also combined demographic information with satellite

imagery data to test whether adjustment for covariates can improve prediction of mortality from satellite images.

Model Interpretation

The output of an intermediate layer of the network was extracted from each image to form *embeddings* – reduced dimensional representations of the data. These *embeddings* (represented by 1000-dimensional vectors) could demonstrate how overall features of the images cluster and correspond with mortality or other demographics. These extracted image features were used to make UMAP [36] plots to identify clustering by images. We overlaid the actual images themselves for each of the points, then true mortality and other covariates such as education and aging. A model that has learned to recognized key image features would begin to cluster images with similar morphology. To examine specific associations between the pertinent image features (i.e. embeddings), we clustered image features using the Spectral Clustering approach [37] and found 10 clusters of images. We averaged the covariate information associated with each of the images across each grouping to yield characteristic covariate descriptors for each cluster. As a direct means for assessing the relationship between deep learning predicted mortality and the covariates, we regressed the image predicted mortality against each of these covariates.

A linear regression model to predict mortality using county level demographic characteristics was fit to data from both the training and validation sets, weighted by county population size (to reduce the variance of the parameter estimates), and evaluated on the test set as a comparison method versus the satellite image approach. We note that the goal of this study was to provide a benchmark for how well the CNN could predict mortality and to use covariates to help contextualize what the deep learning model is “seeing”.

Shapley additive feature explanations (SHAP) [38] is an analytical technique that explains complex models using a simpler surrogate model for each testing instance. We applied SHAP to the images of the test counties to form pixel-wise associations with increases or reductions in mortality, hotspots in these images denote important mortality-associated objects that the model has learned; the technique was similarly applied to identify important covariate predictors for each county (Figure 1d). Finally, we convolved learned CNN image filters over select images to further demonstrate which features of the built environment were utilized by deep learning model.

Results

The training, validation and test counties included 1,721,052 deaths per 217,938,597 individuals in the 2015 U.S. population, a crude mortality rate of 7.90 deaths per 1,000 U.S. residents (Supplementary Table 1).

Our deep learning model was able to accurately predict mortality, with a Pearson correlation coefficient of 0.72 (R^2=51%) between the predicted and true mortality (Figure 2a) on a held-out test set.

The linear regression model using county demographics also accurately predicted mortality; the R^2 was 90% (0.95 correlation) between covariates and the mortality rate (Figure 2b). We observed a reduction in 13 deaths per 100,000 individuals for an 1% increase in the proportion of those who attend college (regression coefficient $\beta = -12.6 \pm 0.8$), and increase of 214 deaths per 100,000 individuals for 5-year increases in average population age ($\beta = 2.1 \pm 0.1$); these were the most important predictors of county mortality (Figure 3a-c, Supplementary Table 2-5). Increased Hispanic ($\beta = -6.6 \pm 0.5$), female ($\beta = 42.5 \pm 6.4$), or Asian race proportions

($\beta = -7.8 \pm 1.0$), and living in the western United States ($\beta = -1.1 \pm 0.3$) as compared to the other regions, were found to be protective county-level factors against mortality (Figure 3a-b, Supplementary Tables 2-4).

Neural networks compress high-dimensional image data into lower dimensional representations in the process of making a prediction. We extracted representations/features for each image and relate the images to each other by their features. We visualized these features using a 2-D UMAP plot. We grouped the images into ten clusters and found variation between clusters (Figure 3c-d, 4a). Weighted pairwise t-tests on mean differences between select covariates between all pairs of clusters demonstrated that image features also associate with covariate mortality predictors (Figures 4b, Supplementary Table 6). Images in cluster 7 were associated with the highest mortality rate of all of the clusters (mortality rate of 11 deaths per 1,000 individuals) and generally included counties from the southeastern United States (Supplementary Figure 5). Income and educational status were lower on average compared to the other clusters. We contrast cluster 7 with cluster 2, a small cluster with the third lowest mortality rates (6.6 deaths per 1,000 individuals) of the 10 clusters (Figure 4; Supplementary Figures 6-8; Supplementary Table 6) and identified variation within cluster 5; we include a brief discussion in the supplementary materials. A univariable regression of the deep learning predicted mortality to each of the mortality predictors across the test set demonstrated that these image features are strongly associated with the mortality predictors (Supplementary Table 7).

Standardized regression coefficients from the covariate model and SHAP values over the test counties for the covariate model informed on the importance of each predictor (Figure 3a-b, Supplementary Tables 2-5). When we used SHAP to identify relevant features in satellite images across these test counties, we noticed that the model was able to associate common

features associated with the socioeconomic status of that community to reductions in mortality. Generally, we were able to spot instances in which sidewalks, driveways, curved roads, hiking trails, baseball fields and light-colored roofs were associated with reductions in mortality (Figure 1d, 5). Conversely, we noticed instances where centerlines of large roads and shadows of buildings and trees were positively associated with mortality. To corroborate the evidence found using SHAP, we plotted small images representing patterns that the first convolutional layer had learned and then convolved each of these patterns with three select images (Supplementary Figures 1-4). Across three images from California, Virginia, and Georgia, filters such as 12, 29 and 43 focused on bright cues and were able to highlight driveways, sidewalks, walkways and baseball fields, while filters 13 and 46 were able to pick up on contours and nuanced shape-based patterns.

Sensitivity analyses over the included number of schools and area around schools indicated that mortality predictive performance tends to increase with the number of schools and area sampled around each school. However, predictive performance appeared to level-off after selection of 75 images per county (3 schools per county; 25 images per school; each image occupies $\frac{1}{49}$ square mile) (Supplementary Figure 9). Our covariate adjusted deep learning model appeared to improve mortality predictive performance (Test Pearson $r=0.83$, $p<0.001$; Validation Pearson $r=0.87$, $p<0.001$) but was unable to surpass the performance achieved using only the demographic characteristics (Supplementary Figure 10).

Discussion

Here, we demonstrated the feasibility of using deep learning and satellite imagery to predict county mortality across the United States, extending methods and approaches from prior deep learning studies that linked the built environment and health-related factors. Furthermore, we

established clusters of images that represented various demographic groups and interrogate these learned patterns to find additional corroborating evidence.

While the covariate model significantly outperformed our deep learning model and identified important predictors that corroborated with prior literature, the deep learning approach identified meaningful built features of the environment, representing a benchmark for performance from which to compare future applications[39, 40].

We only sampled neighborhood characteristics found around schools. For instance, there is likely a greater number of sports fields and playgrounds surrounding schools than in random locations in the county. In many of the neighborhoods, regardless of mortality rates, there is likely more green space surrounding schools. While prior literature suggests green space is associated with positive health factors [41], the fact that this green space is nearly ubiquitous in neighborhoods near these schools may cause our model to down-weight these urban design factors. It also appeared that predictive performance increased by sampling the surrounding neighborhood around each school, suggesting that the surrounding neighborhood contains more important information associated with mortality than utilizing the schools alone.

The evidence of what our deep learning model found to be indicative of mortality can be corroborated with existing literature on associations of these image features with higher socioeconomic status, decreased obesity, and greater designs in urban planning [41–43]. For instance, we found that sidewalks, driveways, curvy roads, hiking trails and baseball fields, amongst other factors, were related to lower mortality; these factors were also uncovered from inspection of learned convolutional filters (Supplementary Figures 1-4).

The assigned importance given to image features such as baseball fields does not imply that these components are necessary for urban design but does further elucidate a suite of

modifiable community factors to design interventions. For instance, accessibility to trails, while indicative of a county with high socioeconomic status[44, 45], provides a convenient means to exercise but can also provide access to other portions of a community, allowing more social mobility and access to health care[46, 47].

Despite our strong findings, there were some limitations to our study. The design of this study was ecological with sampling at one time point for both images and mortality (cross-sectional). This precluded our ability to show temporality and that the built environment was causally associated with mortality. The images also were taken in 2018 while we assessed mortality in 2015. While we do not expect areas surrounding schools to change substantially, we cannot confirm the extent of land-use alterations between 2015 and 2018. We also acknowledge the shortfalls associated with utilizing counties as the primary spatial unit of analysis in this study. While county-level demographic and mortality information is more widely and freely available, mortality rates and demographic factors can vary widely within counties; thus, counties may not always represent the best unit of analysis for capturing the complete heterogeneity in population demographics and mortality rates. In future work, longitudinal image monitoring could be used to forecast increases in mortality rate and help determine optimal intervention strategies that complement county/state planning and development efforts. We assessed county mortality rates using a small percentage of schools from the county, making generalizations to counties included in the study and counties not in our dataset. We were able to predict crude mortality rates accurately within a subsample of these counties. However, our primary modeling approach could not delineate how much the corresponding model-identifiable neighborhood features were intermediates reflective of mortality associated demographics or were directly associated with mortality. Additionally, the deep learning model that combined county-level demographic and satellite imagery data did not surpass the performance of the demographic-only model. We incorporated county-level demographic data into images that have varying degrees of importance for

the prediction of mortality; we suspect that adjusting and including the images that are less predictive of mortality may partially explain the performance of the demographic plus satellite image model.

Higher correspondence between predicted and true mortality may be achieved by increasing the number of schools selected but tapers off with a larger sampling area, which is also suggestive that satellite images beyond a certain distance from a school are not as predictive of mortality. This, ultimately, gives further validity to the approach of choosing a school as a POI. In addition, we may be able to identify more ubiquitous points of interest other than schools from which to sample.

While we acknowledge prior literature documenting the potential for shadow effects to confound aerial imagery analysis, we also note that our model was able to pick up on key factors associated with mortality by utilizing learned filters with shape, color and intensity[48, 49] (Supplementary Figures 1-4). Possible removal or augmentation of these shadows may cause the model to focus on other important characteristics pertinent to higher mortality prediction [48–50]. Finally, although we sampled counties randomly, our sampling scheme preferred populous counties, and had assumed that these counties would solely contain suburban and residential land-use patterns. However, even populous counties contained rural areas that may have obfuscated our ability to sample ubiquitous land-use regions for the mortality prediction potentially biasing the results. Recent deep learning works have focused on grasping health factors such as access to care in the rural setting[51] and could be employed in the context of mortality in the future.

The deep learning approaches used were able to achieve remarkable performance given technical challenges associated with sampling images over large geographic regions. Nor were covariate measures or temporality incorporated into the model. Different sampling techniques

[52], feature aggregation measures, evaluating or producing higher resolution images [53], direct estimation of neighborhood demographic factors, and segmentation of various land-use objects can potentially provide more accurate and interpretable models for studying health and disease [24, 54, 55]. The images may have also been sampled during various points throughout the year, thus results may have been affected by seasonality, however, it is likely that the images had been collected randomly with respect to seasonality and the fact that the model was able to distinguish mortality rates further attests of the ability to learn features less tied to seasonality. Future applications could include deep learning methodologies that are able to account for effects of seasonality[56, 57]. Transfer learning from other GIS-corroborated data may also improve the model's performance. Regardless, our approach is scalable and uses open access data enabling further exploration.

This approach to assess mortality using CNNs has much external applicability. Offshoots of this approach may further explore how the built environment affects mortality in more precise estimates, such as cities, or explore rural areas exclusively[58]. Additionally, our approach demonstrates that built image features of the environment are correlated with demographic characteristics. Future mortality research that lacks the ability to attain particular covariate information due to feasibility or expense costs, could thus infer mortality using images alone or in combination with some known covariates. Applications beyond assessing mortality as an outcome could benefit from this approach and could use other landmarks as a sampling strategy for prediction instead of schools. Sampling from landmarks such as schools reduces the stochastic selection of images from large areas and obviate the need to select and download huge swaths of images. However, if the researcher is able to overcome the resources required to download all of the images from within each county, future methodological advances should consider methods to assign importance scores to each satellite image or weight each county by their perceived macro-level importance for improved mortality prediction. Combining

such methods with demographic information and contextualizing satellite images by neighboring images may yield models that ignore these demographic community factors. Ultimately, these tools could then be used by epidemiologists and policy makers to identify clusters of exposures or diseases such as cancer, arsenic exposure, or infectious diseases for more targeted interventions[24].

Conclusion

We found that mortality could be predicted from satellite imagery. Future use of deep learning and satellite imagery may assist in forming targeted public health interventions and policy changes.

Author Contributions:

Concept and design: JJL, RML, and TAM

Acquisition, analysis, or interpretation of the data: All authors

Drafting of the manuscript: JJL, RML, and TAM

Critical revision to the manuscript for important intellectual content: All authors

Statistical analysis: JJL, RML, TAM

Administrative, technical, or material support: JJL, RML and TAM

Supervision: TAM

Funding: This work was supported by NIH grant R01CA216265. JL is supported through the Burroughs Wellcome Fund Big Data in the Life Sciences at Dartmouth. RML was funded under NIAID T32AI007519.

Acknowledgements: The authors would like to thank Christopher L. Henry and Jorge F. Lima for discussion of image interpretations.

Conflicts of Interest: The authors report no conflicts of interest.

Competing Financial Interests: The authors have no competing financial interests.

Availability of data and materials: Mortality death certificate data can be downloaded from CDC Wonder, per third-party request, available at the following URL: https://www.cdc.gov/nchs/data_access/cmf.htm. Google Maps Satellite Images are publicly available and can be downloaded using the Google Static Maps API.

Ethics approval and consent to participate: Not applicable.

Consent for Publication: Not applicable.

References

1. Dyer O. US life expectancy falls for third year in a row. BMJ. 2018;:k5118.

2. Hamidi S, Ewing R, Tatalovich Z, Grace JB, Berrigan D. Associations between Urban Sprawl and Life Expectancy in the United States. International Journal of Environmental Research and Public Health. 2018;15:861.

3. Muennig PA, Reynolds M, Fink DS, Zafari Z, Geronimus AT. America's Declining Well-Being, Health, and Life Expectancy: Not Just a White Problem. Am J Public Health. 2018;108:1626–31.

4. Case A, Deaton A. Mortality and Morbidity in the 21st Century. Brookings Papers on Economic Activity. 2017;2017:397–476.

5. Woolf SH, Schoomaker H. Life Expectancy and Mortality Rates in the United States, 1959-2017. JAMA. 2019;322:1996–2016.

6. Backlund E, Sorlie PD, Johnson NJ. The shape of the relationship between income and mortality in the United States. Evidence from the National Longitudinal Mortality Study. Ann Epidemiol. 1996;6:12–20; discussion 21-22.

7. The Association Between Income and Life Expectancy in the United States, 2001-2014 | Health Disparities | JAMA | JAMA Network. https://jamanetwork.com/journals/jama/article-abstract/2513561. Accessed 29 Oct 2019.

8. Woolhandler S, Himmelstein DU. The Relationship of Health Insurance and Mortality: Is Lack of Insurance Deadly? Ann Intern Med. 2017;167:424.

9. Carter BD, Abnet CC, Feskanich D, Freedman ND, Hartge P, Lewis CE, et al. Smoking and Mortality — Beyond Established Causes. New England Journal of Medicine. 2015;372:631–40.

10. Xu H, Cupples LA, Stokes A, Liu C-T. Association of Obesity With Mortality Over 24 Years of Weight History: Findings From the Framingham Heart Study. JAMA Netw Open. 2018;1:e184587–e184587.

11. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century | PNAS. https://www.pnas.org/content/112/49/15078. Accessed 29 Oct 2019.

12. Jensen JR, Cowen DC. Remote Sensing of Urban/Suburban Infrastructure and Socio-Economic Attributes. In: The Map Reader. John Wiley & Sons, Ltd; 2011. p. 153–63. doi:10.1002/9780470979587.ch22.

13. Lo CP, Faber BJ. Integration of landsat thematic mapper and census data for quality of life assessment. Remote Sensing of Environment. 1997;62:143–57.

14. Tapiador FJ, Avelar S, Tavares-Corrêa C, Zah R. Deriving fine-scale socioeconomic information of urban areas using very high-resolution satellite imagery. International Journal of Remote Sensing. 2011;32:6437–56.

15. Maharana A, Nsoesie EO. Use of Deep Learning to Examine the Association of the Built Environment With Prevalence of Neighborhood Adult Obesity. JAMA Netw Open. 2018;1:e181535–e181535.

16. Nguyen QC, Sajjadi M, McCullough M, Pham M, Nguyen TT, Yu W, et al. Neighbourhood looking glass: 360º automated characterisation of the built environment for neighbourhood effects research. J Epidemiol Community Health. 2018;72:260–6.

17. Suel E, Polak JW, Bennett JE, Ezzati M. Measuring social, environmental and health inequalities using deep learning and street imagery. Sci Rep. 2019;9:1–10.

18. Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S. Combining satellite imagery and machine learning to predict poverty. Science. 2016;353:790–4.

19. Tingzon I, Orden A, Sy S, Sekara V, Weber I, Fatehkia M, et al. Mapping Poverty in the Philippines Using Machine Learning, Satellite Imagery, and Crowd-sourced Geospatial Information. :5.

20. Keralis JM, Javanmardi M, Khanna S, Dwivedi P, Huang D, Tasdizen T, et al. Health and the built environment in United States cities: measuring associations using Google Street View-derived indicators of the built environment. BMC Public Health. 2020;20:215.

21. Nguyen QC, Khanna S, Dwivedi P, Huang D, Huang Y, Tasdizen T, et al. Using Google Street View to examine associations between built environment characteristics and U.S. health outcomes. Preventive Medicine Reports. 2019;14:100859.

22. Robinson C, Hohman F, Dilkina B. A Deep Learning Approach for Population Estimation from Satellite Imagery. arXiv:170809086 [cs]. 2017. http://arxiv.org/abs/1708.09086. Accessed 29 Oct 2019.

23. Gebru T, Krause J, Wang Y, Chen D, Deng J, Aiden EL, et al. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. PNAS. 2017;114:13108–13.

24. Weichenthal S, Hatzopoulou M, Brauer M. A picture tells a thousand…exposures: Opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology. Environment International. 2019;122:3–10.

25. Multiple Cause of Death Data on CDC WONDER. https://wonder.cdc.gov/mcd.html. Accessed 29 Oct 2019.

26. USDA ERS - County-level Data Sets. https://www.ers.usda.gov/data-products/county-level-data-sets/. Accessed 22 Nov 2019.

27. Personal Income by County, Metro, and Other Areas | U.S. Bureau of Economic Analysis (BEA). https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas. Accessed 22 Nov 2019.

28. 19 Age Groups - Standard Populations - SEER Datasets. SEER. https://seer.cancer.gov/stdpopulations/stdpop.19ages.html. Accessed 29 Oct 2019.

29. Geographic. https://nces.ed.gov/programs/edge/Geographic/SchoolLocations. Accessed 29 Oct 2019.

30. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014;12:1495–9.

31. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. Journal of The Royal Society Interface. 2018;15:20170387.

32. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44.

33. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. Advances in Neural Information Processing Systems 25. Curran Associates, Inc.; 2012. p. 1097–1105. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf. Accessed 9 Jun 2019.

34. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv:151203385 [cs]. 2015. http://arxiv.org/abs/1512.03385. Accessed 30 Jul 2019.

35. Deng J, Dong W, Socher R, Li L-J, Li K, Li FF. ImageNet: a Large-Scale Hierarchical Image Database. 2009. p. 248–55.

36. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:180203426 [cs, stat]. 2018. http://arxiv.org/abs/1802.03426. Accessed 5 Mar 2019.

37. Belkin M, Niyogi P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Computation. 2003;15:1373–96.

38. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems 30. Curran Associates, Inc.; 2017. p. 4765–4774.

http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf. Accessed 9 Jun 2019.

39. Muller A. Education, income inequality, and mortality: a multiple regression analysis. BMJ. 2002;324:23.

40. Montez JK, Zajacova A. Trends in Mortality Risk by Education Level and Cause of Death Among US White Women From 1986 to 2006. Am J Public Health. 2013;103:473–9.

41. Gelormino E, Melis G, Marietta C, Costa G. From built environment to health inequalities: An explanatory framework based on evidence. Preventive Medicine Reports. 2015;2:737–45.

42. Aneshensel CS, Harig F, Wight RG. Chapter 15 - Aging, Neighborhoods, and the Built Environment. In: George LK, Ferraro KF, editors. Handbook of Aging and the Social Sciences (Eighth Edition). San Diego: Academic Press; 2016. p. 315–35. doi:10.1016/B978-0-12-417235-7.00015-9.

43. Rutt CD, Coleman KJ. Examining the relationships among built environment, physical activity, and body mass index in El Paso, TX. Preventive Medicine. 2005;40:831–41.

44. Wilson DK, Kirtland KA, Ainsworth BE, Addy CL. Socioeconomic status and perceptions of access and safety for physical activity. Ann Behav Med. 2004;28:20–8.

45. Engelberg JK, Conway TL, Geremia C, Cain KL, Saelens BE, Glanz K, et al. Socioeconomic and race/ethnic disparities in observed park quality. BMC Public Health. 2016;16:395.

46. Moore LV, Diez Roux AV, Evenson KR, McGinn AP, Brines SJ. Availability of Recreational Resources in Minority and Low Socioeconomic Status Areas. Am J Prev Med. 2008;34:16–22.

47. Thornton RLJ, Glover CM, Cené CW, Glik DC, Henderson JA, Williams DR. Evaluating Strategies For Reducing Health Disparities By Addressing The Social Determinants Of Health. Health Aff (Millwood). 2016;35:1416–23.

48. Kwatra V, Mei Han, Shengyang Dai. Shadow removal for aerial imagery by information theoretic intrinsic image analysis. In: 2012 IEEE International Conference on Computational Photography (ICCP). Seattle, WA, USA: IEEE; 2012. p. 1–8. doi:10.1109/ICCPhot.2012.6215222.

49. Corke P, Paul R, Churchill W, Newman P. Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2013. p. 2085–92.

50. Bansal N, Akashdeep, Aggarwal N. Deep Learning Based Shadow Detection in Images. In: Krishna CR, Dutta M, Kumar R, editors. Proceedings of 2nd International Conference on Communication, Computing and Networking. Singapore: Springer; 2019. p. 375–82.

51. Satellite images and machine learning can identify remote communities to facilitate access to health services | Journal of the American Medical Informatics Association | Oxford Academic. https://academic.oup.com/jamia/article/26/8-9/806/5549818. Accessed 29 Oct 2019.

52. Jean N, Wang S, Samar A, Azzari G, Lobell DB, Ermon S. Tile2Vec: Unsupervised representation learning for spatially distributed data. In: AAAI. 2018.

53. Shermeyer J, Van Etten A. The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery. arXiv:181204098 [cs]. 2019. http://arxiv.org/abs/1812.04098. Accessed 29 Nov 2019.

54. Bischke B, Helber P, Folz J, Borth D, Dengel A. Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks. In: 2019 IEEE International Conference on Image Processing (ICIP). 2019. p. 1480–4.

55. Iglovikov V, Mushinskiy S, Osin V. Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition. arXiv:170606169 [cs]. 2017. http://arxiv.org/abs/1706.06169. Accessed 29 Nov 2019.

56. Fu H, Gong M, Wang C, Batmanghelich K, Zhang K, Tao D. Geometry-Consistent Generative Adversarial Networks for One-Sided Unsupervised Domain Mapping. :10.

57. Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXiv:170310593 [cs]. 2018. http://arxiv.org/abs/1703.10593. Accessed 29 Nov 2019.

58. Bruzelius E, Le M, Kenny A, Downey J, Danieletto M, Baum A, et al. Satellite images and machine learning can identify remote communities to facilitate access to health services. J Am Med Inform Assoc. 2019;26:806–12.

Figures

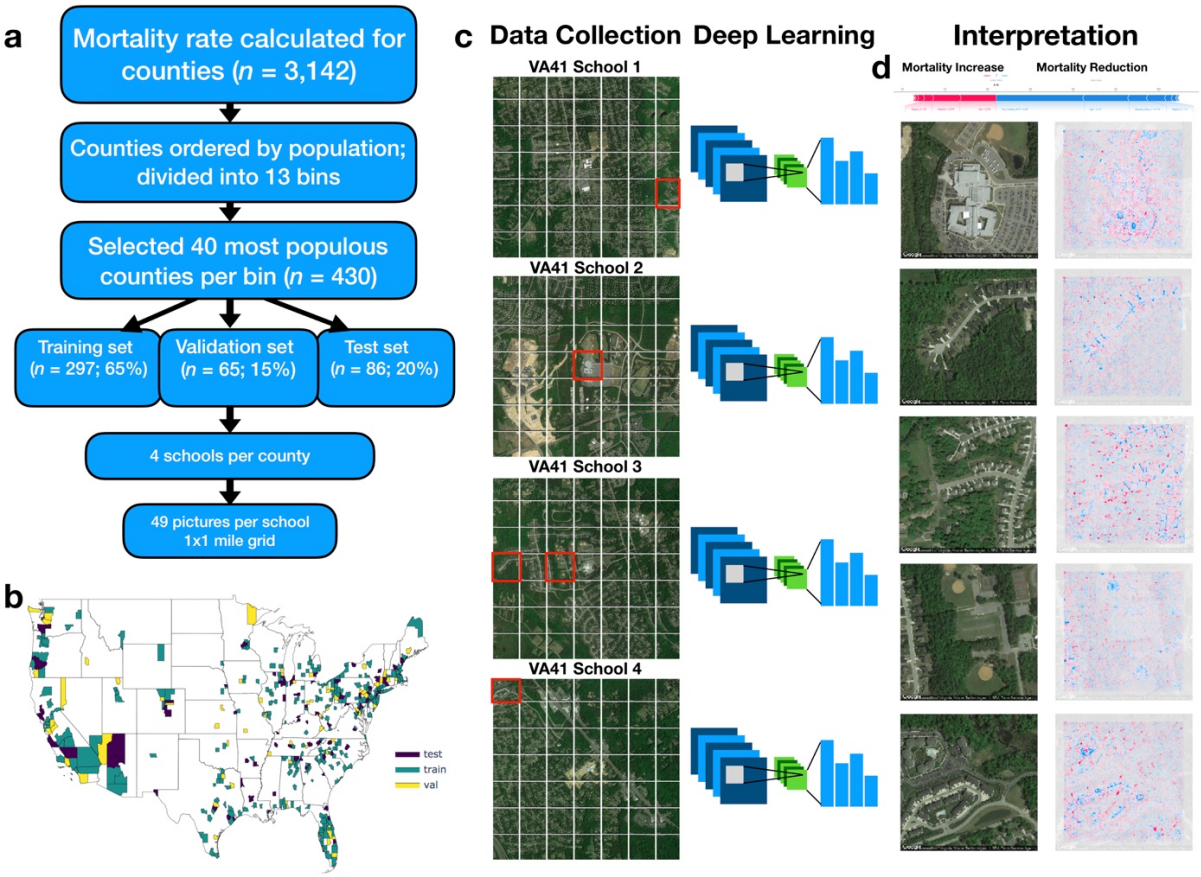


Figure 1: Method Overview: a) Flow diagram depicting selection of counties and sampling scheme for images; 13 mortality bins created from 1000 most populous counties to capture diversity of mortality rates; b) Map of the United States colored by selection of training, testing and validation counties; c) 4 sets of 1x1 mile grid of images (7 images by 7 images) were selected per each county, all images randomized and used to train deep learning model; visual depiction of deep learning: filters slide across each image and pick up on key image features, information is aggregated to form model predictions; the predictions for each of these 196 images per county are averaged to yield final mortality rates; d) image features interrogated using SHAP and compared to important covariates from linear model for each test county.

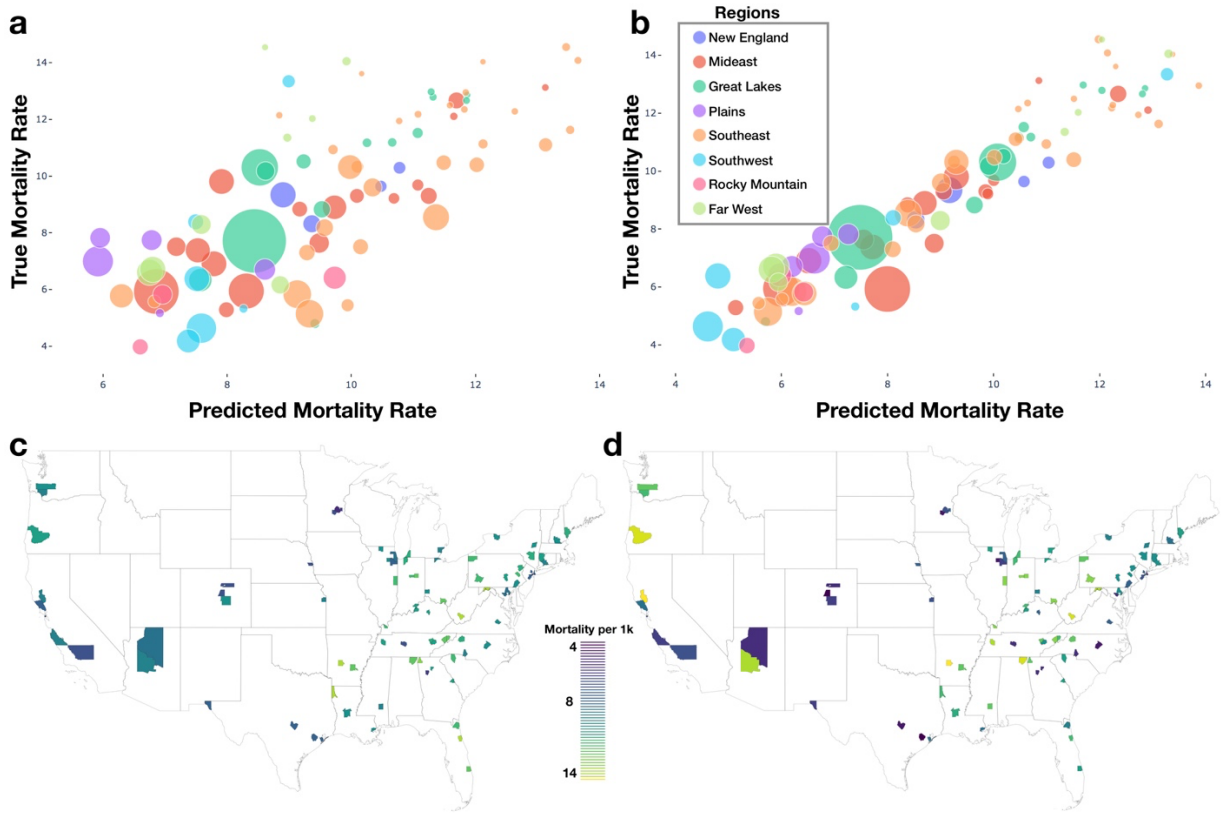


Figure 2: Model Results: a) deep learning predicted versus true mortality for each test county; b) linear model predicted versus true mortality for each test county; a-b) legend contains look-up dictionary for naming of US regions featured in the rest of the study; size of bubble is correspondent to population size of county; c) predicted mortality plotted geographically for test counties; d) true mortality plotted geographically for test counties

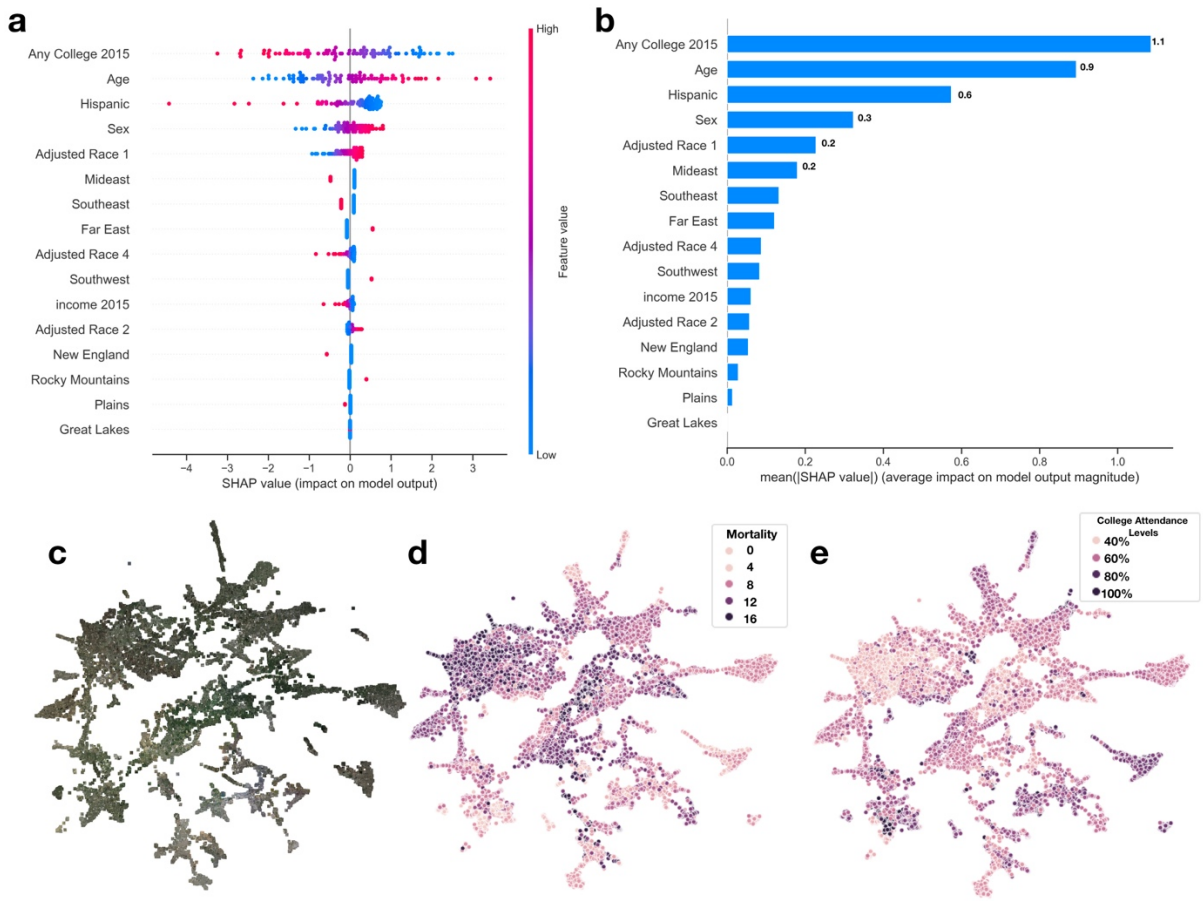
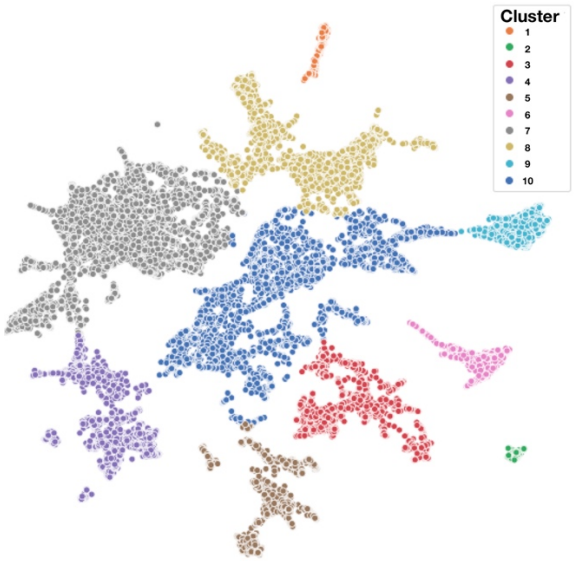


Figure 3: SHAP Summary of Covariate Predictors and Image Embeddings a) SHAP summary plot for covariate model evaluated on test set, covariates ranked by feature importance; b) SHAP rankings of overall importance of covariates; c) overall covariate importance over test set; d-e) 2-D UMAP plot of image features derived from images using deep learning model: d) actual images overlaying 2-D coordinates; e) colored by true county mortality; f) colored by average college attendance; we note here that our model has been explicitly trained to associate image features with mortality; the separability of college attendance over the image features is an artifact of the association between education status and mortality

a



b

Cluster	Predicted Mortality	True Mortality	Sex	Hispanic	Race White	Race Black	Race Asian
1	7.4	6.2	0.5	0.1	0.7	0.2	0.1
2	6.8	6.6	0.5	0.1	0.6	0.1	0.4
3	9.4	8.3	0.5	0.1	0.7	0.2	0.1
4	7.5	7.0	0.5	0.2	0.8	0.1	0.1
5	8.3	8.5	0.5	0.3	0.9	0.0	0.0
6	6.6	6.5	0.5	0.1	0.8	0.1	0.1
7	11.0	11.0	0.5	0.1	0.9	0.1	0.0
8	10.0	9.5	0.5	0.1	0.8	0.2	0.0
9	7.3	7.0	0.5	0.2	0.6	0.3	0.1
10	10.3	10.2	0.5	0.1	0.8	0.1	0.0

Cluster	Population	College Attendence (%)	Income 2015 (USD)	Age	Absolute Residual	Number Images
1	595 000	71.4	50 000	7.9	1.3	269
2	844 000	75.4	106 000	8.7	0.2	117
3	776 000	66.1	59 000	8.2	1.8	1502
4	677 000	64.6	59 000	7.9	1.4	1652
5	456 000	61.7	44 000	8.4	2.8	1205
6	731 000	73.1	62 000	8.0	1.8	691
7	206 000	54.9	43 000	8.6	1.3	4290
8	666 000	60.1	45 000	8.4	1.7	2465
9	1 138 000	58.0	52 000	8.2	0.6	686
10	275 000	58.0	45 000	8.5	1.7	3979

Figure 4: Key Covariate Characteristics of Found Image Clusters: a) 2-D UMAP plots colored by cluster; b) averaged covariate statistics for each image cluster; average population and income estimates rounded to thousands place

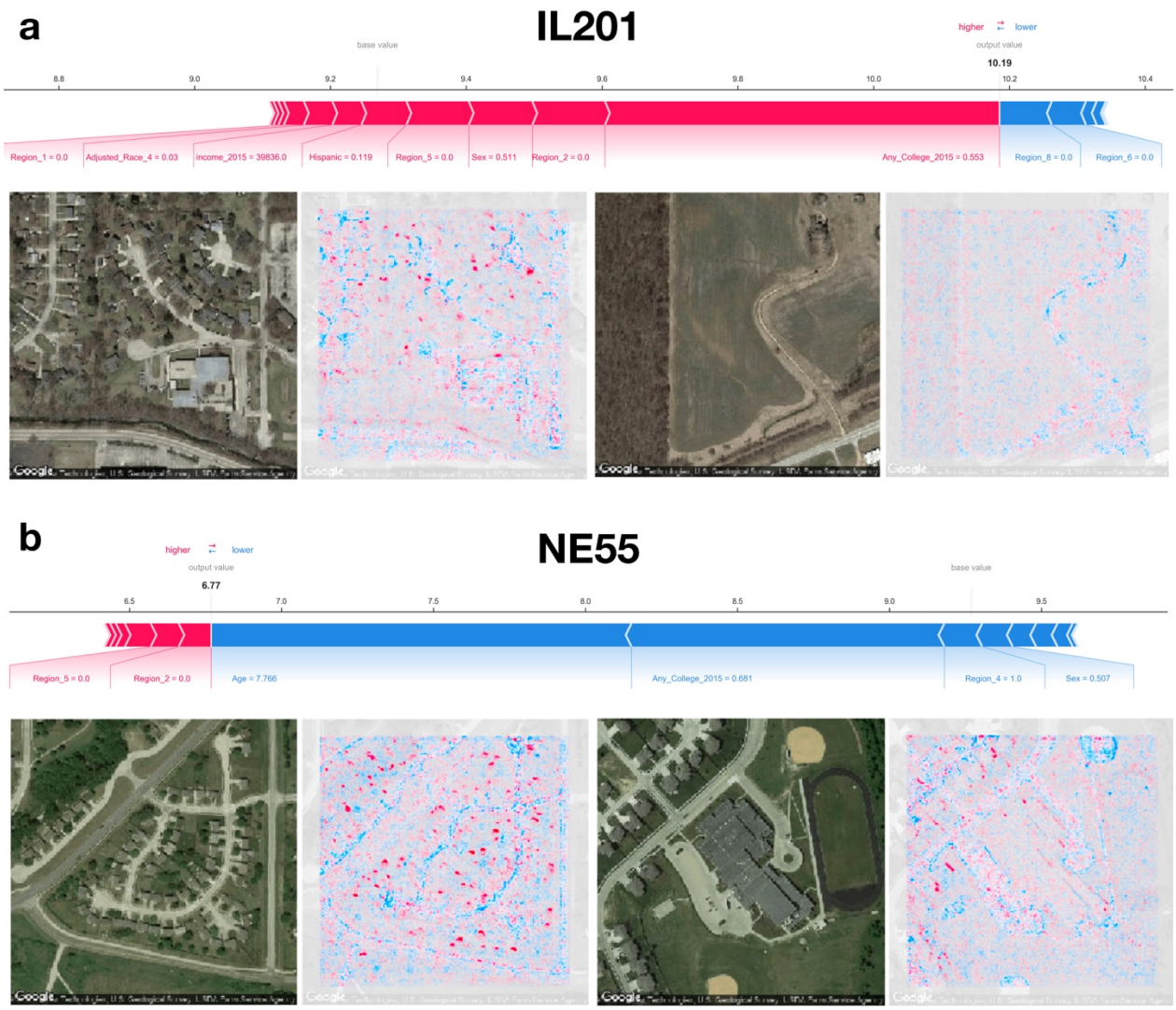


Figure 5: SHAP Image Interpretations: a-b) Shapley features extracted from both covariate and image models for select images from counties in Illinois and Nebraska (FIPS codes IL201 and NE55 respectively); blue coloring indicates features associated with reductions in mortality, red indicates association with increased mortality; additional image interpretations can be found in Figure 1d