Mental Health and Land Cover: A Global Analysis Based on Random Forests

Abstract

Nature features and processes in living environments can help to reduce stress and improve mental. Different land types have disproportionate impacts on mental health. However, the relationships between mental health and land cover are inconclusive. Here, we show the complex relationships between mental health and percentages of eight land types based on the random forest method. The accuracy of our model is 90.51%, while it is normally no more than 20% in the previous studies. According to the analysis results, we estimate the average monetary values of eight land types, which are an 8.24%-GDP-per-capita (95% confidence interval: 8.17% - 8.32%) increase in the individual income for wetland, 0.44% (0.39% - 0.49%) for bare land, 0.31% (0.39% - 0.49%) for shrubland, 0.10% (0.06% - 0.14%) for forest, 0.09% (0.05% - 0.12%) for water, 0.04% (0 % - 0.07%) for cropland, -0.21% (-0.25% - -0.18%) for urban land, and -0.35% (-0.39 % - -0.31%) for grassland. Due to complex relationships, the current land cover composition influences people's attitudes towards a certain land type. This paper proves that the relationships between mental health and land cover are complex and far from linear. Furthermore, it provides insights to formulate better land-use policies.

Keywords:

Mental Health; Land Cover; Environmental Valuation; Machine Learning; Random Forest

Introduction

Natural land cover in people's living environments has positive impacts on human well-being and mental health (Alcock et al., 2014; Bratman et al., 2019; Diaz et al., 2019; Diener et al., 2018; MacKerron and Mourato, 2013; Malek and Verburg, 2020), which is mainly driven by ecosystem service (Costanza et al., 1997; Costanza et al., 2014; Diaz et al., 2018). For example, environments with natural land cover provide a variety of ecosystem services that can benefit people (Carpenter et al., 2009; Chaplin-Kramer et al., 2019; Diaz et al., 2018; Diaz et al., 2019), such as recreational activities (MacKerron and Mourato, 2013), air pollution abatement (Chameides et al., 1988; Eitelberg et al., 2016; Mendoza-Ponce et al., 2018), and the creation of aesthetic, artistic, and scientific values for human beings (Felipe-Lucia et al., 2018; Seresinhe et al., 2015). However, 2.7% of global semi-natural or natural land has been continuously converted to other land types, specifically cropland and built-up area, from 1992 to 2015 (OECD, 2017). Due to losing natural land cover, the estimated aggregate value of ecosystem services from 1997 to 2011 has decreased by $4.3 trillion globally and annually (Costanza et al., 2014).

As the benefits of the natural land cover are vast and enormous (Costanza et al., 2014; Diaz et al., 2018), there is a need to clarify the effect of land cover change on mental health. Because the world continues to develop and urbanize (OECD, 2017), the share of natural land cover in people's living environments will likely keep decreasing. Due to the trade-off between development and the desire for natural land, there is a further need to detect how much alteration of land cover composition affects future mental health.

Land cover data are various in previous studies, which are share of one or several land types in each defined area (Alcock et al., 2014; Kopmann and Rehdanz, 2013; Krekel

et al., 2016; Tsurumi et al., 2018; White et al., 2013), or the land types where participants are in (MacKerron and Mourato, 2013; Seresinhe et al., 2019). When the researchers use the retrospective assessments of SWB, normally, they take the share of land types as the land cover data (Kopmann and Rehdanz, 2013; Krekel et al., 2016; Tsurumi et al., 2018; Tsurumi and Managi, 2015). People's preference for a certain land type and its effects on mental health can be analyzed. Based on the relationship among mental health or well-being, the income of respondents and land cover in their living environments, monetary values of land covers are calculated (Tsurumi et al., 2018; Tsurumi and Managi, 2015, 2017). Substantial and significant evidence shows that people in natural environments experience higher well-being and better emotions (MacKerron and Mourato, 2013; Seresinhe et al., 2019). These researches provide justifications that natural environments positively affect human well-being and mental health (MacKerron and Mourato, 2013; White et al., 2013). Indicators of land cover data are the percentages of several land types in some articles (Alcock et al., 2015; Kopmann and Rehdanz, 2013; Krekel et al., 2016; Tsurumi et al., 2018), while the share of greenery is the only land cover variable is employed in other articles (Alcock et al., 2014; Tsurumi and Managi, 2015). To investigate the relationships between mental health and each land type, this study uses the percentage of eight land types as the land cover data.

Previous studies have explicitly focused on seeking to express the effects of land cover in monetary terms (Barrio and Loureiro, 2010; Saphores and Li, 2012; Tsurumi and Managi, 2015), because this will efficiently deliver the economic loss from devastated land cover to both the public and policymakers (Li and Managi, 2021; Managi and Kumar, 2018). In this study, we choose to express the value in the monetary term as in previous

studies (Krekel et al., 2016; Tsurumi et al., 2018; Tsurumi and Managi, 2015). The widely used method first estimates the marginal effects of non-market goods and income. Then it uses the ratio of those marginal effects to indicate the monetary value of non-market goods (Frey et al., 2010). However, this method only works when the relationships are linear or similar to linear. For example, in the previous studies, some studies assume that the coverage rate of a certain land type is linearly related to well-being or mental health (Alcock et al., 2015; Alcock et al., 2014; Ambrey and Fleming, 2014; Tsurumi et al., 2018), while in others the relationship are considered non-linear, either quadratic (Krekel et al., 2016; Tsurumi and Managi, 2015) or logarithmic (Kopmann and Rehdanz, 2013). However, the relationships estimated by machine learning methods might be more complex than the abovementioned assumptions. Therefore, further processes of machine learning results are needed in our study.

The advantage of the machine learning methods is high accuracy. The goodness of fit in previous studies is no more than 20%. The relationship between mental health and land cover is mainly assumed linear (Alcock et al., 2015; Tsurumi et al., 2018; White et al., 2013), quadratic, or logarithmic (Kopmann and Rehdanz, 2013; Krekel et al., 2016; Tsurumi and Managi, 2015). On the one hand, the linear relationship is straightforward and has a clear attitude towards a certain land type. These models are based on a simple assumption that amounts of certain land types always have the same effect on mental health, no matter the current land cover status. In this case, people should live in the environment only with this land type that has the most positive effect on mental health. This is the main shortcoming of this assumption and is far from reality. On the other hand, the non-linear relationship is more in line with reality. Preferences for certain land types depend on the

current land cover allocation (Kopmann and Rehdanz, 2013; Krekel et al., 2016; Tsurumi and Managi, 2015). It is the fundamental idea to build a non-linear model. There are two types of non-linear models. One assumes the relationship between the coverage of land types and well-being is logarithmic (Kopmann and Rehdanz, 2013), and the other assumes the relationship is quadratic (Krekel et al., 2016; Tsurumi and Managi, 2015). In the logarithmic relationship assumption, when the certain coverage keeps increasing, the effect of this land type on well-being or mental health gets down, but the direction of this attitude does not change (Kopmann and Rehdanz, 2013). In the quadratic relationship assumption, when the share of land cover changes, the intensity of effects on mental health will vary and even may alter the direction of the impact. Though these non-linear assumptions are better than linear ones, they still have low accuracy. The machine learning method's accuracy typically exceeds 80%. High accuracy means that the relationships estimated by the trained model are closer to the actual situation. To make the policies based on the analysis results reliable, we must guarantee assumptions similar to the real world. Machine learning has fewer assumptions compared with previous methods. Therefore, using machine learning methods to analyze the relationships is necessary.

Materials and Methodology

Materials

Survey Information

Our study is based on an international survey conducted by Kyushu University from July 2015 to March 2017, covering 37 countries, including both developed countries and

developing countries. The investigation periods for each country are generally less than one month. Moreover, to guarantee the reliability of the survey, the same questionnaires are used, while currency-related questions are based on local currencies. The population and GDP of these countries account for 68.58% of the global population and 82.67% of the worldwide GDP in 2017, respectively (**Appendix Table A1**). This survey was conducted to obtain individual mental health and several other demographic and socio-economic characteristics. The total number of observations that were recorded is 100,956. However, due to a lack of geographical location or records, 95,571 observations are kept. In addition, because some individuals did not provide income information, 88,730 observations are used in the calculations (Descriptive statistics of the features shown in **Appendix Table A2**).

Mental Health

To access individual mental health, we include the twelve-item General Health Questionnaire (GHQ-12) in the survey. As a validated and reliable instrument, the GHQ-12 is widely used in previous mental health studies, e.g. (El-Metwally et al., 2018; Hankins, 2008; Quek et al., 2001). The GHQ-12 comprises 12 items to assess the individual mood states. Each item of the GHQ-12 has four potential answer options, and they are encoded by the Likert method as 0, 1, 2, and 3, from negative to positive. The mental health assessment score is computed as the summed score of all 12 items. Thus, the output variable of our study is a discrete numeric variable ranging from 0 to 36. The current random forest method is designed to execute either regression or classification. Using the discrete output variable, the algorithm would perform the classification task by assuming

the output is a categorical variable. However, adjacent scores of the mental health assessments are related, i.e., they are ordinal rather than categorical. **Figure 1** illustrates the statistical distribution of the mental health assessment score. Most people get 24 points in the assessment, and significantly more people score 24 to 30 points than others. In this situation, if we perform the classification random forests, the classification accuracy for the people with lower or higher scores would be extremely low due to the unbalanced output distribution. Thus, we assume that the mental health assessment score is continuous.

Global Land Cover Data

As for the land cover, we use remote sensing data compiled by Tsinghua University (http://data.ess.tsinghua.edu.cn/), because, to the best of our knowledge, it is the data set with the highest global resolution, at approximately 30 meters. This data set is the 2017 global land cover. It classifies land cover into ten categories: cropland, forest, grassland, shrubland, wetland, water, tundra, urban land, bare land, and snow/ice (Gong et al., 2019). We calculated areas of each land type surrounding each respondent of our survey with these data. To estimate the impact of land cover in our analysis, we use the percentages of each land type within a radius of 5,000 meters around each respondent following the previous study (Krekel et al., 2016). Only eight land types are taken as the land cover data because little tundra and snow/ice are in the analyzed area. After the random forest analysis, we will build partial dependency profiles of land cover variables. A certain land type range is divided into many small intervals. If there are many intervals without any respondents, the partial dependency profiles are unreliable. Therefore, to improve the estimation accuracy, we reduce the ranges of the percentages of shrubland, wetland, water, and bare

land. The range of shrubland is set from 0 to 40% because only 282 respondents live with more than 40% shrubland and some even exceed 60%. Similarly, the wetland, water, and bare land ranges are set from 0 to 3%, from 0 to 50%, and 0 to 20%, respectively.

Individual Income Data

Converting the impacts of land cover on mental health into monetary values is an effective way to make them understandable to the public without professional knowledge. Moreover, individual income is also an essential factor affecting mental health. In this study, we use the difference between individual income and GDP per capita in the respondent's country (DIG) as the income variable because the income in the survey is based on local currency. Additionally, the same amounts of money have different effects in different countries. For example, the ability and sense of 100 USD in the U.S. and Sri Lanka are not actually the same for the local people. The DIG is calculated as follows:

$$DIG_i = \frac{Inc_i - GDPPC_i}{GDPPC_j} \quad (1)$$

where DIG_i is the DIG of the respondent i, Inc_i is the individual income of the respondent i, and $GDPPC_j$ is the GDP per capita of the respondent i's country in the surveyed year. It is must be noted that the units of Inc_i and $GDPPC_i$ are the current price USD. In the survey, the income data is based on the local currencies. To unify the income data, we convert the currencies into USD. We employed the official annual average exchange rate of the year during the conversion process when the survey was conducted in that country. Moreover, the survey questionnaires required the respondents to select their gross household income range rather than to ask them to report their exact gross household income. Thus, we take

the midpoint of the range selected by the respondent as the gross household income. For instance, in the U.S., if respondents reported that their gross household income ranged from 50,000 to 60,000 USD per year, their household income is 55,000 USD per year in the analysis. According to previous studies (Mackerron and Mourato, 2009; Yuan et al., 2018), the calculation of the annual gross individual income is as follows:

$$Inc_i = \frac{GHI_i}{(Adu_i + 0.7Chi_i)^{0.5}} \quad (2)$$

where Inc_i is the annual gross individual income of the respondent i, GHI_i represents the gross annual household income of the respondent i, Adu_i represents the number of adults in the respondent i's household, and Chi_i represents the number of children in the respondent i's household. Limited by the household size and the maximum value of the selection interval, individual income rarely exceeds three times GDP per capita in the respondent's country. To improve the estimation accuracy, we set the range of the difference from -1 to 3. **Figure 2** demonstrates the statistical distribution of the DIG in the respondent's country.

Other Control Variables

We add several other control variables because mental health status may differ according to people's socio-economic and demographic characteristics, which are age, gender, employment, education background, emotion in the surveyed week, children number, self-reported health, self-reported personality, and evaluation of living environment. Among these control variables, employment, education background, and self-reported personality are categorical. We use the one-hot encoding method to convert

them into a series of dummy variables. Thus, in the analysis, every respondent has 47 features and one output variable. The descriptions of the features are listed in **Appendix Table A3**.

Methodology

To detect influential factors on mental health and confirm the relationship between mental health and land cover, linear regression methods, such as ordinary least square (OLS) and ordered logit regression (OLR), is widely applied, e.g., (Akpinar et al., 2016; Alcock et al., 2015; Alcock et al., 2014; Li and Managi, 2021). The studies evaluate the monetary values of land cover through OLS estimation because the OLS is straightforward to explain. Additionally, the investigations employing the OLR are theoretically more reasonable since mental health evaluation is a discrete variable rather than a quantitative and continuous variable in most studies (Alcock et al., 2015; Alcock et al., 2014). The OLR is a typical classification function based on logistic regression. However, these two models rely on linear assumptions and cannot directly illustrate the importance of predictors on the outcome variable. Putting another way, based on the linear assumption, a 1-unit increase in a certain land type always has the same effect on mental health, whatever the status quo. This is not consistent with the actual situation. Random forest could smoothly grasp the non-linear relationship, which bundles many different decision tree algorithms as an improved boosting method (Breiman, 2001). The decision tree algorithms are non-linear algorithms and closer to real-world situations (Czajkowski and Kretowski, 2016). Moreover, according to the out-of-bag (OOB) test, each feature's importance on the output variable is estimated.

The Regression Decision Tree

A single decision tree is the fundamental element of the random forest method. There are two types of trees, namely, the decision tree for either classification or regression (Breiman, 2001; Liaw and Wiener, 2002). **Figure 3** shows a simple example of a regression decision tree with three layers. To complete the prediction in the example tree, the algorithm passes three internodes and does three judgments at most. As the example illustrated in **Figure 3**, we assume only three features, the unemployed dummy variable, self-reported health, and the DIG, affect the output variable, mental health. The rules of each judgment and feature range splits are the critical part of machine learning training. A large amount of data is put into the algorithm to train the decision tree to decide the rules of each judgment and feature range splits. We employ a greedy approach to train regression decision trees (Breiman et al., 2017). This approach chooses the features and splits their ranges to minimize the residual sum of squares (RSS) as follows:

$$RSS = \sum_{l \in leaves} \sum_{i \in C_l} (y_i - \bar{y}_{C_l})^2 \quad (3)$$

where l is a leave, C_l is the cases in leaf l, y_i is the observed value and $\bar{y}_{C_l}$ is the average observed value in leaf l. Unless the RSS is smaller than the defined threshold or the number of remaining cases in the end leaf is less than the defined threshold, the number of internode of trees will keep increasing (Breiman et al., 2017).

Random Forest

In most cases, a single regression decision tree is insufficient to fit the output variables and usually causes an over-fitting analysis. To solve this issue, the random forest ensemble a bundle of decision trees and let them vote for the results (Breiman, 2001). The voting strategy for regression is taking the average of all individual predictions as the random forest prediction. Bagging and bootstrapping are performed to guarantee the accuracy and reliability of the random forest (Liaw and Wiener, 2002). Bootstrapping is the sampling technique of the random forest. Firstly, we set the number of trees in our random forest as N_{tree}. We extract N_{tree} samples with replacement from the original data and the sample sizes are 2/3 data of the total sample. Every decision tree utilizes the bootstrapped data set. However, only N_{try} random features are used in a single decision tree, rather than all. For regression tasks, N_{try} is roughly one third of the total number of features. After training, the random forest can predict the output variable through the aggregate of the votes from each tree. Using the bootstrapped data set and the aggregate of votes, this process is terminologically called "bagging". Additionally, roughly 1/3 of the total sample is not employed in the training process named the OOB data set. The OOB data set is applied to test the accuracy of the random forest through the OOB error rate, which is the proportion of OOB observations incorrectly predicted by the trained random forest. Due to the bagging technique, cross-validation is unnecessary. The reliable trained models have a relatively low OOB error rate.

In the random forest, most parts are built randomly, while only two critical parameters must be decided by the users, specifically, N_{tree} and N_{try}. The accuracy of the random forest will increase to some extent when more trees are included. However, the

cost of infinitely increasing N_{tree} is a dramatic increment of calculation power and calculating time. Moreover, when N_{tree} exceeds a particular value, the marginal effect of increasing the number is minimal. Accordingly, taking the size of our data set and calculation power into account, the number of trees is set to 1,000. Moreover, the number of the features used in the decision trees, N_{try}, is another vital factor. A large N_{try} might lead to overfitting, while a small N_{try} might cause underfitting. Previous studies indicate that one third of the total number is optimal (Breiman, 2001; Breiman et al., 2017; Liaw and Wiener, 2002). Our data set has 47 features, so N_{try} should be 17.

Variable Importance

The random forest could estimate the importance of each feature on the output variable. The basic idea of importance estimation in the random forest is to calculate the increase in mean squared error (IncMSE) between before and after excluding a certain feature (Breiman, 2001). IncMSE of a specific feature would be higher when it is more important to successfully predict the output variable compared with other features. IncMSE is similar to the partial R^2 in the OLS algorithm. There is no need to select the features in the random forest algorithm since the issues, such as multicollinearity, do not influence the accuracy of the random forest. Yet, multicollinearity is a fatal problem in the OLS.

Partial Dependency Profile

Although the accuracy of random forest is high, it is challenging to understand and explain the results (Friedman, 2001; Greenwell, 2017). Partial dependency profiles (PDPs)

offer a simple solution to make the agnostic models human-understandable. PDPs illustrate the relationship between the output variable and features of interest in two dimensions. PDPs are estimated as follows:

$$\widehat{f_A}(A_l) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(A_l, \boldsymbol{OF}_i) \tag{4}$$

where $\widehat{f_A}(A_l)$ is the average predicted outcome given a value A_l of a specific feature of interest A when all other features are actual values from the original data set, A_l is a given feature value belonging to the range of feature A in the original data set, $\boldsymbol{OF}_i$ is a vector of feature values except A of the respondent i, and n is the data set size, 88730, in this study. All A_l from an arithmetic progression (AAP) from the minimum to the maximum of the feature A. Simply speaking, the PDP uses the trained model to predict the output variable with a specific feature value, assuming other feature values are constant. The results of PDP are matrixes including two columns. One column is the feature of interest, and the other is the predicted output variable. If we put the matrixes on a two-dimensional figure, the relationships between the features of interest and the output variable are illustrated. However, generally, they are not smooth splines. In other words, the functions between the features of interest and the output variable are not continuous functions that can be derived everywhere. Thus, the explanation ability of PDPs is still limited.

Pseudo Partial Dependency Function

To solve this issue, we use regressions to build pseudo partial dependency functions (PPDFs) between the features of interest and the output variable. Since the PDPs in this

study are all complex splines, we employ 20-order splines to fit the relationships. The equation is as follows:

$$POV_j = \beta_0 + \beta_1 FOI_j + \beta_2 FOI_j^2 + \cdots + \beta_{20} FOI_j^{20} \quad (4)$$

where POV_j is the predicted output variable of jth row in the PDP matrix, FOI_j is the feature of interest value of jth row, and β_0 to β_{20} are the estimated parameters of the PPDF. The rows in the PDP matrix should be treated differently to make the PPDF close to reality. A specific feature value represents an interval with a width of a common difference of the AAP. Some intervals of specific values cover a large number of respondents, while others might have no respondents within them. Therefore, the intervals covering more respondents should be allocated higher weights to build the PPDF. The vector of the estimated parameters of the PPDF is calculated as follows:

$$\boldsymbol{\beta} = (\boldsymbol{FOI}^T \boldsymbol{W} \boldsymbol{FOI})^{-1} \boldsymbol{FOI}^T \boldsymbol{W} \boldsymbol{POV} \quad (5)$$

where $\boldsymbol{\beta}$ is the vector of the estimated parameters in **Equation 4**, $\boldsymbol{FOI}$ is the matrix of the feature of interest values with 0 to 20 order, $\boldsymbol{POV}$ is the predicted output variable column of the PDP matrix, and $\boldsymbol{W}$ is the weights of each row. In this study, we cannot directly take the respondent counts in each interval as the weights because the counts vary hugely. Using them, the PPDF will have several extrema in a small extent in many cases, which is far from reality. Hence, we employ a function to squeeze the weights' extent as follows:

$$\boldsymbol{W} = \ln(\boldsymbol{Count} + e) \quad (6)$$

where $\boldsymbol{Count}$ is the vector of respondent counts of each interval.

Based on the estimated parameters $\boldsymbol{\beta}$, the PPDF is built and simply written as follows:

$$MH = g_{FI}(FI) \quad (7)$$

where MH is the predicted mental health scores, FI is a certain feature of interest, and g_{FI} is the relationship between mental health and the feature input.

Monetary Values of Land Cover

To make the impacts of land cover change on mental health understandable, we estimate the monetary values of land cover. We take the marginal substitution rate (MSR) of land cover and income as the monetary values. Since the relationship between mental health and land cover are non-linear, a one-unit increase in a certain land type should be clearly defined. In the analysis, we estimate the marginal effects of a one-hectare increase in a certain land type in respondent's living environment on mental health, as follows:

$$ME_{la} = g_{la}(LA_i + \Delta LA) - g_{la}(LA_i) \quad (8)$$

where g_{LA} is the PPDF between mental health and a certain land type la, LA_i is the percentage of land cover la in the living environment of the respondent i, and ΔLA is a one-hectare increase in land type la. The MSR is expressed as follows:

$$g_{la}(LA_i + \Delta LA) - g_{la}(LA_i) = g_{inc}(INC_i + \Delta INC) - g_{inc}(INC_i) \quad (9)$$

where g_{inc} is the PPDF between mental health and the DIG, INC_i is the DIG of the respondent i, and Δinc is monetary value of the land type. The g_{Inc} is not monotonically increasing. Whether the monetary values of a certain land type are negative or positive is decided by the derivation of g_{LA}. If not, some respondents should reduce their income to offset the negative impact of an increase in a land type. This is not consistent with reality. Therefore, we forcedly define the direction of land cover monetary value.

Results

In this study, the trained random forest employs 1000 trees. To train each tree, 17 features are randomly chosen in the bootstrapped data sets. The accuracy of the random forest is 90.51%, whereas the accuracy of the OLS is only 60.30%. Moreover, the root mean square error (RMSE), mean square error (MSE), and mean absolute error (MEA) of the random forest are 1.94, 3.77, and 1.08, respectively, while the RMSE, MSE, and MEA of the OLS are 5.76, 33.22, and 4.54. In terms of accuracy, the random forest in this study significantly exceeds the linear regression. However, the random forest has its own issues. **Figure 4** demonstrates the relationship between predicted and measured mental health scores. The slope of the fit line between predicted and measured mental health scores is lower than 1. The random forest rarely predicts extreme values, such as 0 and 36. Putting another way, the random forest's prediction is closer to the mean value of the output variable. It must be noted that we have already added the weights in the model training process. The weights are the reciprocal of the counts of each possible output value in the total data set. The extreme output values, such as 0 and 36, have obtained more attention in the model training process, though it improves accuracy marginally. In fact, the current model is the best version after several improvements.

Figure 5 demonstrates the importances of each feature. Emotions, including sadness, pleasure and smile, and self-reported health, affect mental health the most. For example, if we do not employ the feature, "Sadness", in the model, the RMSE will increase 4.41. Significantly, the income and land cover in respondents' living environment also

influence mental health. The RMSE will rise 2.58 without the DIG in the model. Moreover, the importance values of grassland, forest, cropland, shrubland, water, urban land, wetland, and bare land, are equal to 2.47, 2.42, 2.40, 2.39, 2.39, 2.38, 2.31, and 2.29 increase in the RMSE, respectively.

Figure 6 illustrates nine PDPs of DIG and land cover features. The relationships between mental health and these variables are non-linear. As shown in **Figure 6.a**, the positive impact of income on mental health is strong, when the income is low. However, the increase in the individual income contributes little to mental health when the income exceeds roughly 1.2 times GDP per capita (the DIG is 0.2). When the income is more than 2.5 times GDP per capita (the DIG is 0.2), the further increase in the income even reduces the mental health score. Previous studies indicate that the turning points in the relationship between emotional well-being and income exist (Diener et al., 2018; Jebb et al., 2018). Once the income fulfills the needs and material desires, the positive impacts of the income on happiness would be limited (Diener et al., 2010). **Figure 6.b - Figure 6.i** demonstrate the relationships between mental health and eight land types. The increases in each land type positively affect mental health when the percentages of land types are close to zero. Moreover, except urban land and grassland, the percentages of other land types in most respondents' living environments are within the increasing interval of the PDPs. The spline of the urban land's PDP is relatively flat, from 25% to 85%, and then rapidly descends. These indicate that people prefer to live in environments with several land types.

We employ 20-order splines to fit the PDPs. The accuracies of PPDFs are 99.79% (DIG), 99.71% (bare land), 99.68% (cropland), 99.62% (forest), 99.60% (grassland), 99.70% (urban land), 98.92% (shrubland), 99.82% (water), and 97.75 (wetland), respectively, as

shown in **Figure 7**. Among eight land types, the monetary value of wetland is the highest. A 1-hectare increase in living environments equals an 8.24%-GDP-per-capita (95% confidence interval: 8.17% - 8.32%) increase in the individual income. In previous research, the wetlands have high ecological values and provide vital ecosystem service (Costanza et al., 1997; Costanza et al., 2014; Kopmann and Rehdanz, 2013). The monetary value of bare land is 0.44% (95% CI: 0.39% - 0.49%) GDP per capita per hectare. It must be underscored that bare land is not abandoned land. The monetary values of shrubland, forest, water, and cropland are positive, which are 0.31% (95% CI: 0.39% - 0.49%), 0.10% (95% CI: 0.06% - 0.14%), 0.09% (95% CI: 0.05% - 0.12%), and 0.04% (95% CI: 0 % - 0.07%) GDP per capita per hectare, resperively. Urban land's and grassland's monetary values are negative, which are -0.21% (95% CI: -0.25% - -0.18%), and -0.35% (95% CI: -0.39 % - -0.31%) GDP per capita per hectare.

Discussion

Our main findings are that the relationships between mental health and land cover are non-linear and non-monotonous. Increases in each land type positively impact mental health when the percentages of these land types are low. Accordingly, it could be implied that people prefer to live in environments rich in diversity and extremely monolithic landscapes might cause mental disorders. Furthermore, it is the first study that fits PDPs to PPDFs. Because PDPs are not always able to be derived everywhere, they only provide the shapes of relationships. The shapes are not enough to estimate the MSR of two features of interest. Thus, we replace PDPs with PPDFs in the analysis, though it losses a little

accuracy. Based on the PPDFs, the monetary values of a one-hectare increase in each land type are estimated, which are an 8.24%-GDP-per-capita (95% CI: 8.17% - 8.32%) increase in the individual income for wetland, 0.44% (95% CI: 0.39% - 0.49%) for bare land, 0.31% (95% CI: 0.39% - 0.49%) for shrubland, 0.10% (95% CI: 0.06% - 0.14%) for forest, 0.09% (95% CI: 0.05% - 0.12%) for water, 0.04% (95% CI: 0 % - 0.07%) for cropland, -0.21% (95% CI: -0.25% - -0.18%) for urban land, and -0.35% (95% CI: -0.39 % - -0.31%) for grassland. Moreover, the model's accuracy is relatively high, indicating the reliability of the results. The accuracy, RMSE, MSE, and MEA are 90.51%, 1.94, 3.77, and 1.08, respectively, exceeding most previous studies.

Previous studies focus more on the impacts of green space on human well-being or mental health in the city (Alcock et al., 2015; Alcock et al., 2014; Krekel et al., 2016; Tsurumi et al., 2018; White et al., 2013). The coverage percentage of green space positively affects mental health (Alcock et al., 2015; Alcock et al., 2014; White et al., 2013). In our study, except grassland and urban land, other land types are positively related to mental health when their percentages are low. These results are consistent with previous studies because they mainly concentrate on the urban area with less coverage of other land types. The impact of grassland is always negative, which is counterintuitive. Two reasons cause this problem. Firstly, this research is based on remote sensing data. In the remote sensing process, grassland is easier to be misclassified especially when close to cropland and shrubland (Gong et al., 2019). Additionally, sporadic grass is more likely to be misclassified, so the low accuracy of grassland in urban areas might mislead the model's results. Secondly, the PDP of grassland shows that the predicted mental health score is an upward trend when grassland coverage is from 0.08% to 0.5%, as shown in **Figure 6.e**.

However, the upward extent is too short, sandwiched between two decreasing intervals. So, when we build the PPDF of the grassland, the function is monotonously decreasing. Though we used several approaches, such as changing the weight squeezing function, reducing the order of spline, among others, the current version is the best, in terms of the accuracy and the line shape. Wetland is the most preferred, as it provides the most amount of ecosystem service (Costanza et al., 1997; Costanza et al., 2014), and it is scarce in the living environment. Bare land's average value is high, but it will negatively affect mental health after the coverage exceeds 1%. According to the figure in the data provider's article (Gong et al., 2019), the large area of bare land is generally dessert, while it in the cities might be sports play yards. Forest and shrubland are similar and positively impact mental health when they are scarce. People cannot enter the large aggregated forest to have various nature experiences, and they are also associated with the possibility of crime (Lee and Maheswaran, 2011; Markevych et al., 2017). The high percentage of urban land is associated with poor mental health. Living in cities naturally is necessary (Douglas et al., 2017; Hartig and Kahn, 2016; White et al., 2017). The adverse effects of large amounts of non-urban land types on mental health indicate that people living in rural areas are likely to have mental disorders and need more assistants.

There are several limitations and issues worthy of note. First, the land cover variables are the percentages of 8 land types in the buffers with a 5-km radius surrounding the living locations of respondents. There is an assumption that the quality of land cover does not influence the effects of those land types on mental health. For example, there may be no difference between a well-designed urban park and grassland in the pasture. Furthermore, the impacts of the distance to a certain land type are ignored. Secondly, this

study only uses the global cross-sectional data, so it cannot detect the difference within people when land cover changes. Global research using panel data to probe the effects within individuals is still desired. Finally, the number of respondents in each country is not the same or proportional to the country's population. The countries with more respondents have more substantial impacts on the results. Thus, the results might be prejudiced, though this database is one of the biggest databases in this field. In future studies, the long panel data should be used to investigate the impacts of land cover within indiviudals. Moreover, more geographical characters should be taken into account. Effective explaining methods and tools should be developed to make the machine learning results understandable.

Conclusion

The relationships between land cover in living environments and mental health are more complex than linear assumptions. An unsuitable increase in a certain land type might not make residents mentally healthy. Based on the current situation, the average monetary values of a one-hectare increase in each land type are estimated, which are an 8.24%-GDP-per-capita (95% CI: 8.17% - 8.32%) increase in the individual income for wetland, 0.44% (95% CI: 0.39% - 0.49%) for bare land, 0.31% (95% CI: 0.39% - 0.49%) for shrubland, 0.10% (95% CI: 0.06% - 0.14%) for forest, 0.09% (95% CI: 0.05% - 0.12%) for water, 0.04% (95% CI: 0 % - 0.07%) for cropland, -0.21% (95% CI: -0.25% - -0.18%) for urban land, and -0.35% (95% CI: -0.39 % - -0.31%) for grassland. Our study illustrates the heterogeneity of the impacts of eight land types on mental health to provide more

information for governments and the public to formulate land-use policies to improve mental health.

Data Availability

The fully reproducible code are publicly available at https://github.com/MichaelChaoLi-cpu/MentalHealthAndLandCover . Data are available from the corresponding author on reasonable request.

Acknowledgement

This research was supported by the following funding agencies: JSPS KAKENHI (Grant No. JP20H00648), the Environment Research and Technology Development Fund of the Environmental Restoration and Conservation Agency of Japan (Grant No. JPMEERF20201001), and also JST SPRING (Grant No. JPMJSP2136).

Figure:

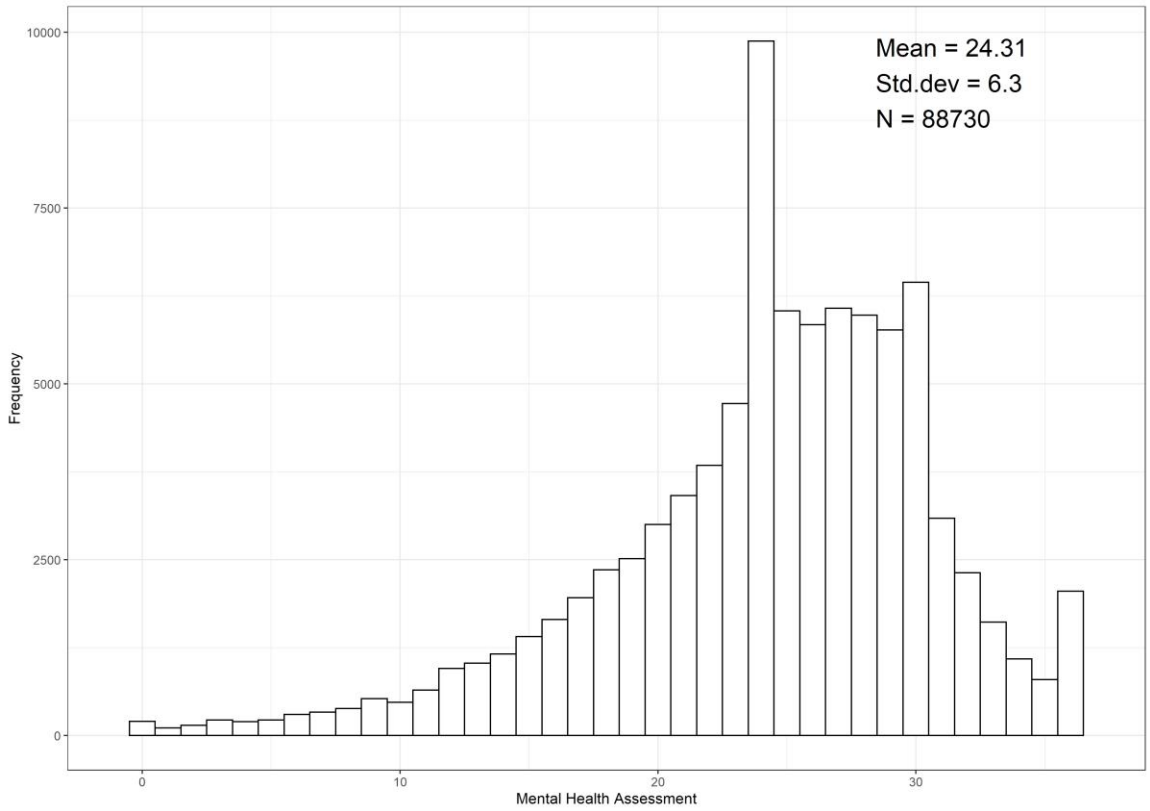

Figure 1: The Statistical Distribution of Mental Health Assessment

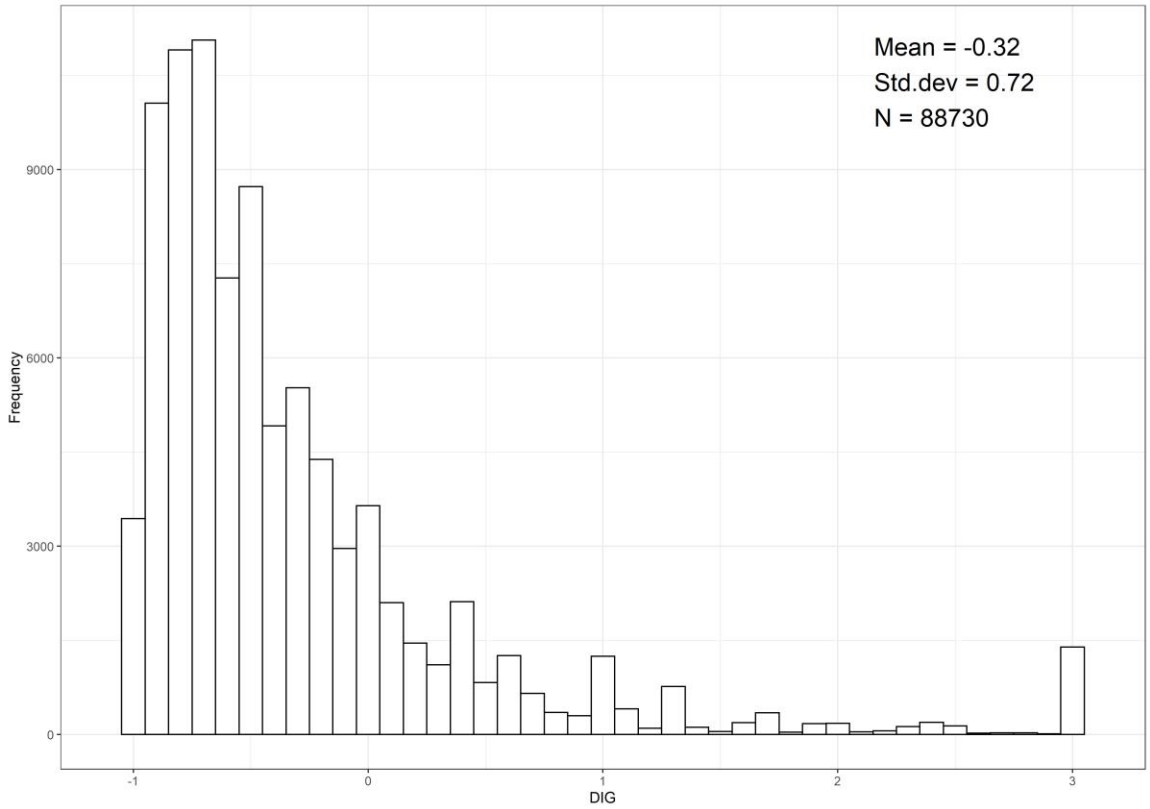

Figure 2: The Statistical Distribution of DIG

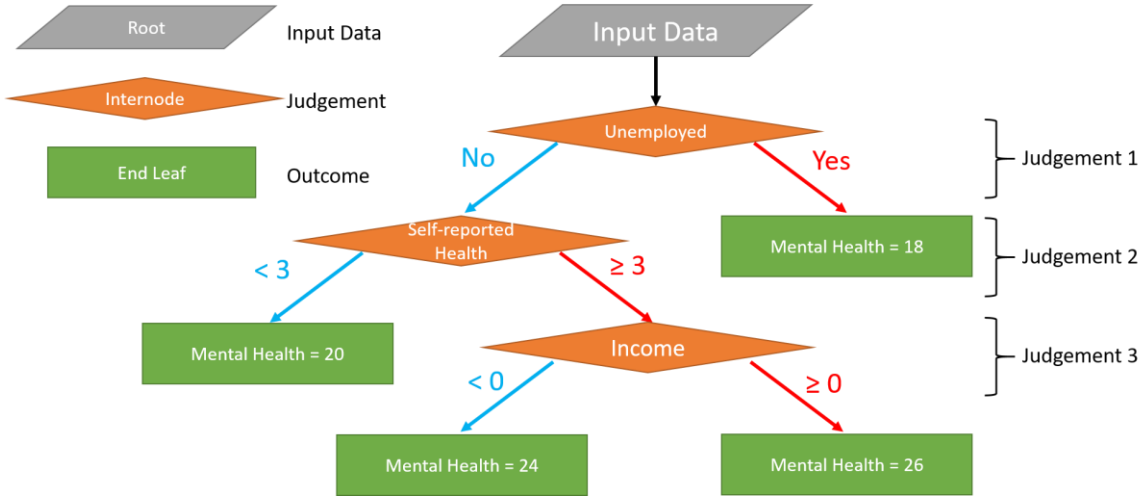

Figure 3: Example of a Regression Decision Tree

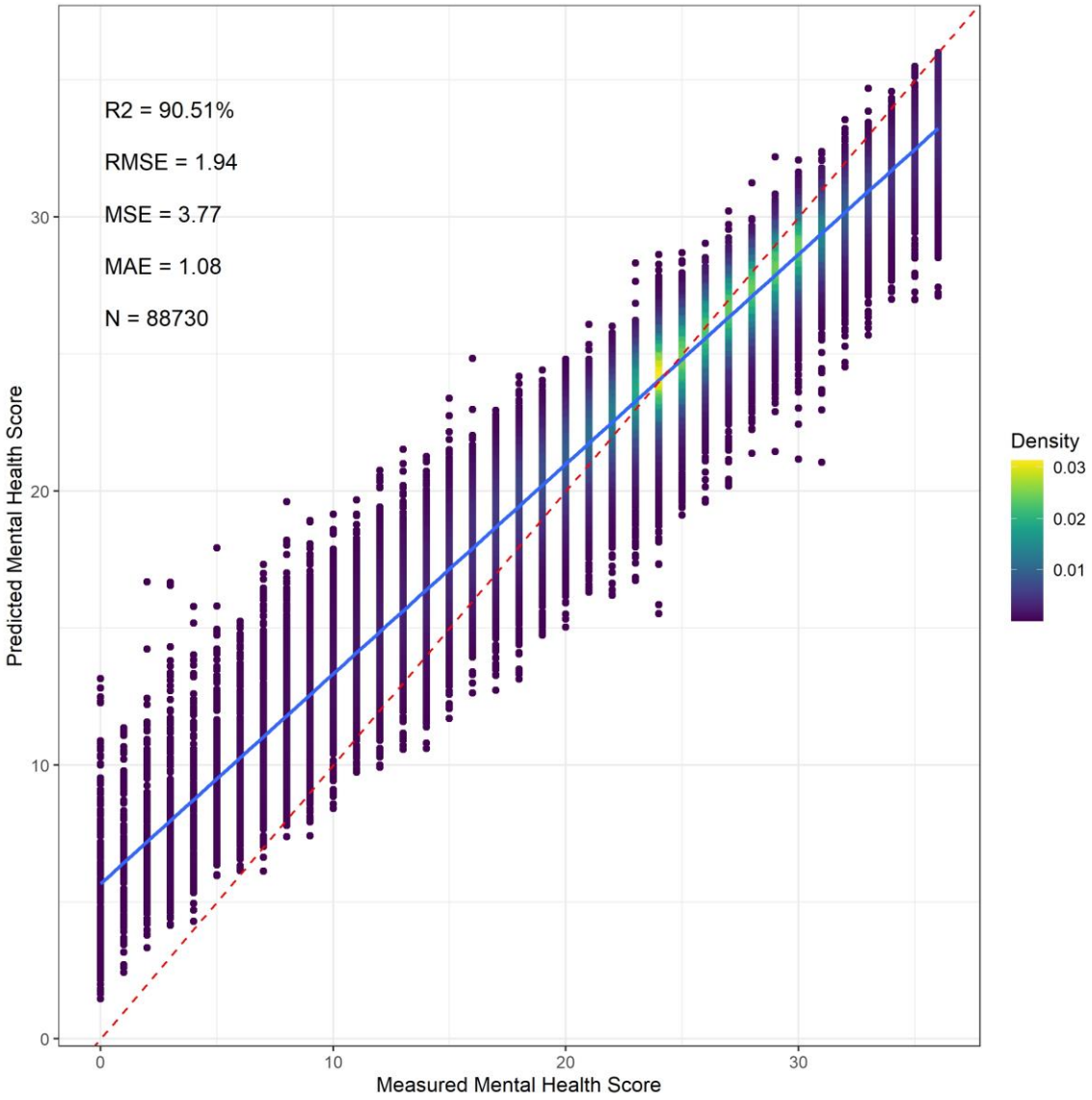

Figure 4: The Density Plots between the Measured and Predicted Mental Health Score

(The red dashed line is the 1:1 line. The blue line is the regression line.)

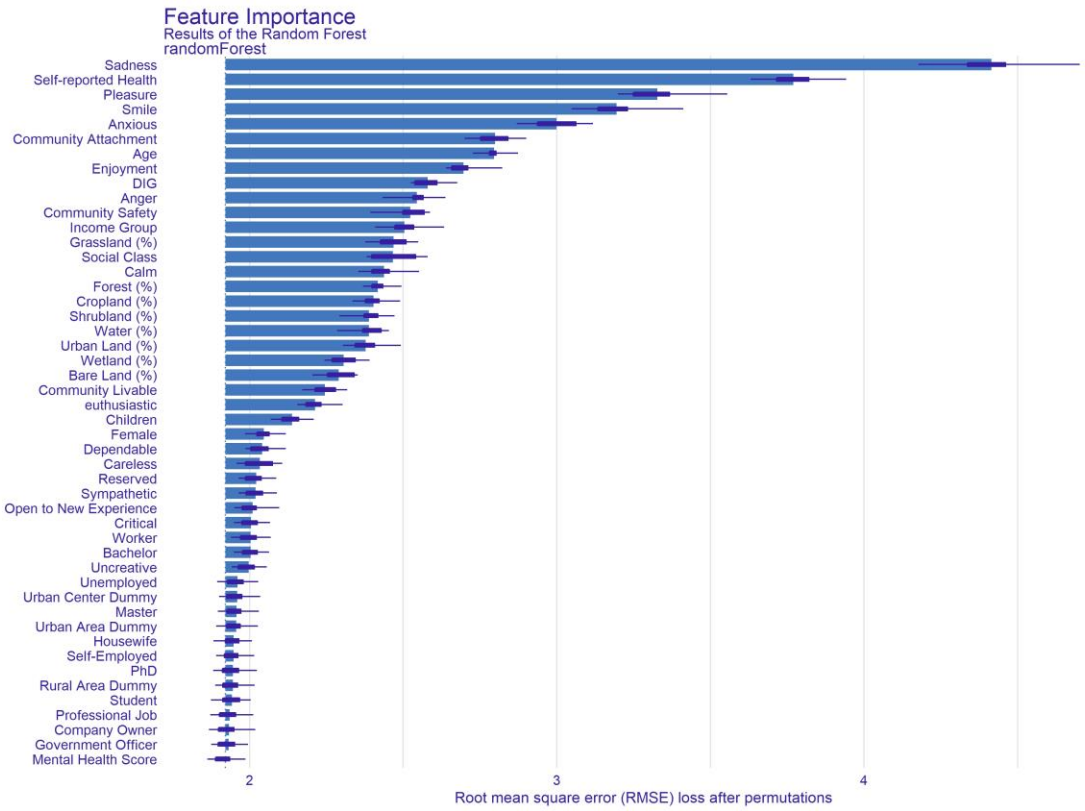

Figure 5: The Importance of Features

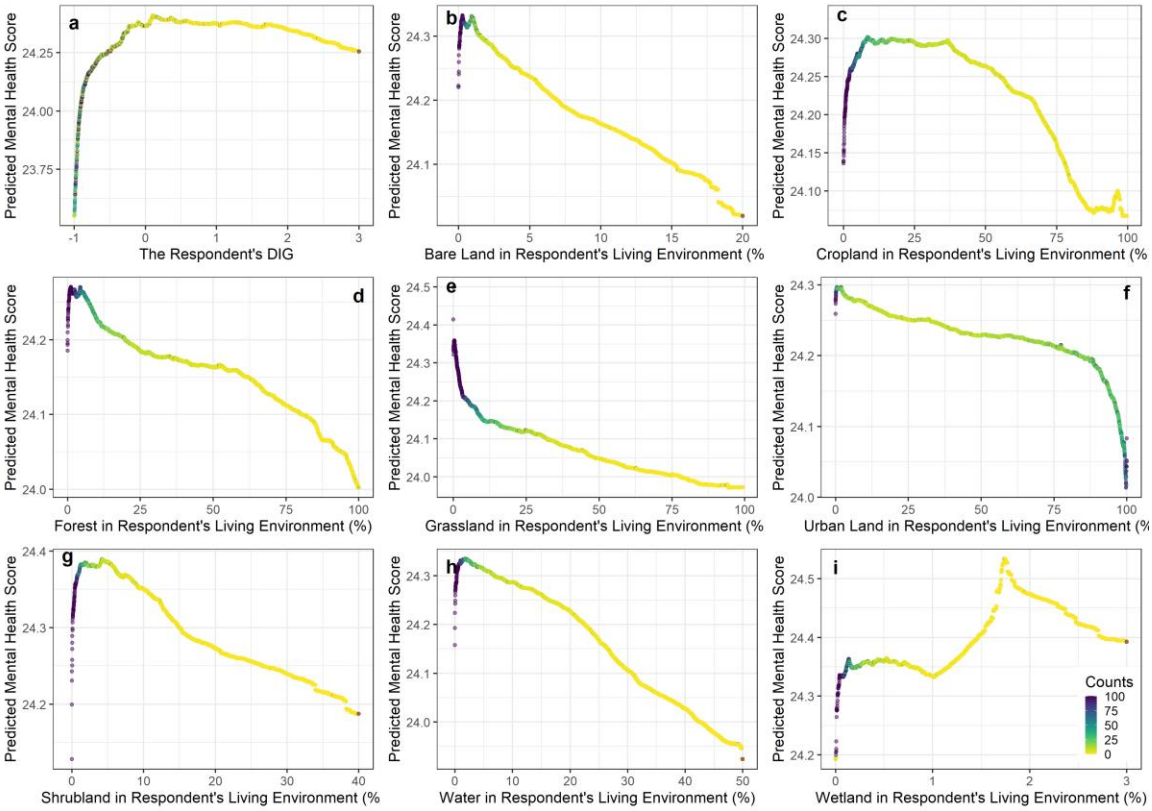

Figure 6: The PDPs of the DIG and Land Cover Features

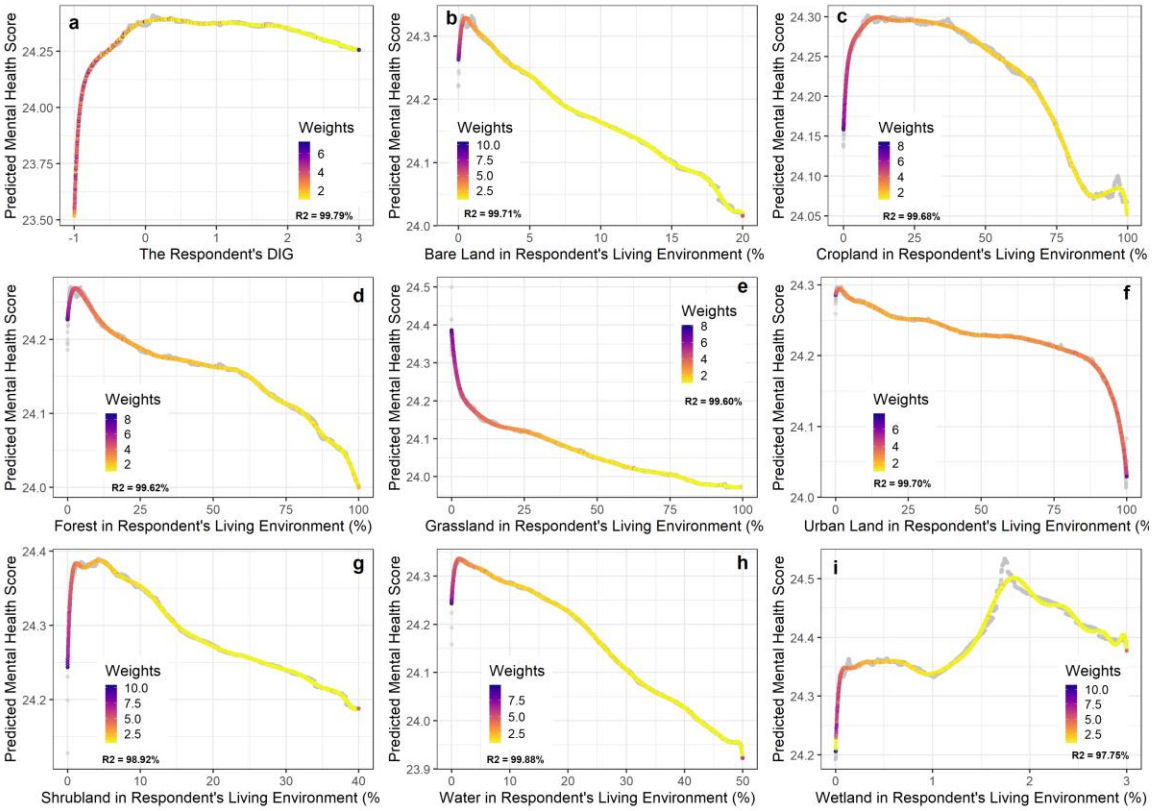

Figure 7: The PPDFs of the DIG and Land Cover Features

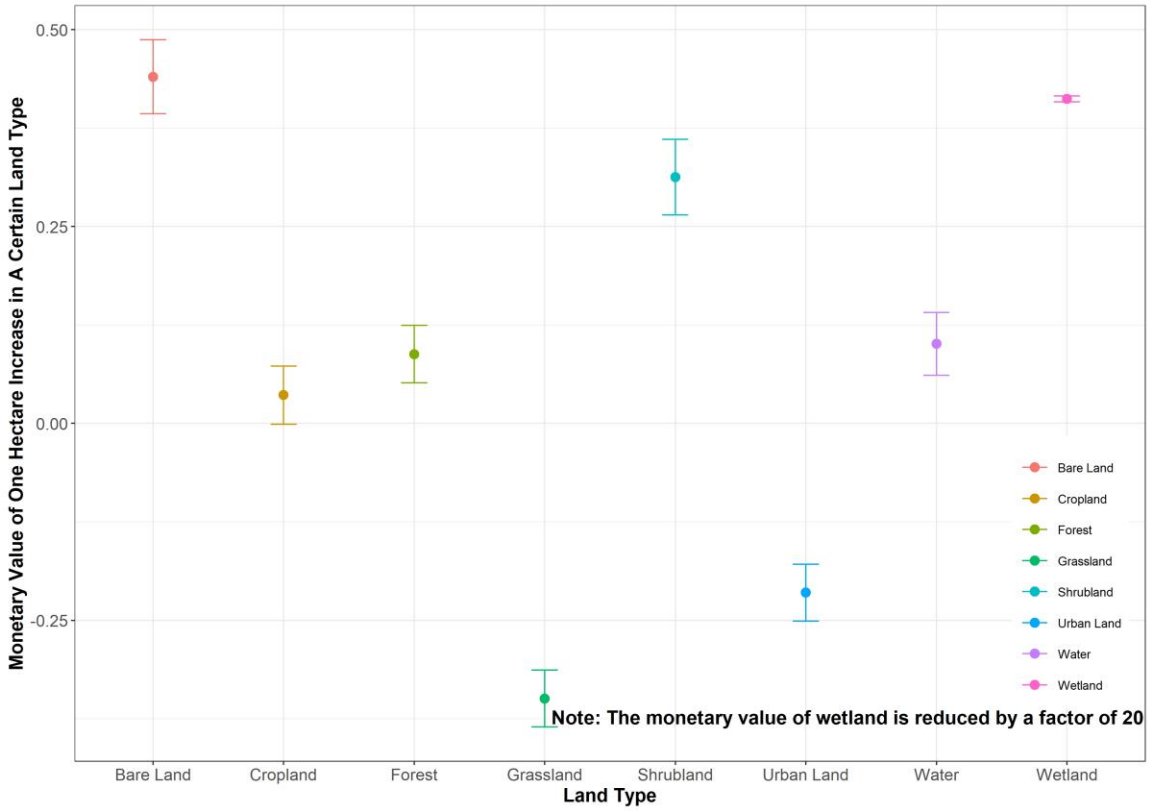

Figure 8: The Monetary Values of One Hectare

Reference:

Akpinar, A., Barbosa-Leiker, C., Brooks, K.R., 2016. Does green space matter? Exploring relationships between green space type and health indicators. Urban Forestry & Urban Greening 20, 407-418.

Alcock, I., White, M.P., Lovell, R., Higgins, S.L., Osborne, N.J., Husk, K., Wheeler, B.W., 2015. What accounts for 'England's green and pleasant land'? A panel data analysis of mental health and land cover types in rural England. Landscape and Urban Planning 142, 38-46.

Alcock, I., White, M.P., Wheeler, B.W., Fleming, L.E., Depledge, M.H., 2014. Longitudinal Effects on Mental Health of Moving to Greener and Less Green Urban Areas. Environmental Science & Technology 48, 1247-1255.

Ambrey, C., Fleming, C., 2014. Public Greenspace and Life Satisfaction in Urban Australia. Urban Studies 51, 1290-1321.

Barrio, M., Loureiro, M.L., 2010. A meta-analysis of contingent valuation forest studies. Ecological Economics 69, 1023-1030.

Bratman, G.N., Anderson, C.B., Berman, M.G., Cochran, B., de Vries, S., Flanders, J., Folke, C., Frumkin, H., Gross, J.J., Hartig, T., Kahn, P.H., Kuo, M., Lawler, J.J., Levin, P.S., Lindahl, T., Meyer-Lindenberg, A., Mitchell, R., Ouyang, Z.Y., Roe, J., Scarlett, L., Smith, J.R., van den Bosch, M., Wheeler, B.W., White, M.P., Zheng, H., Daily, G.C., 2019. Nature and mental health: An ecosystem service perspective. Science Advances 5, eaax0903.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5-32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification And Regression Trees.

Carpenter, S.R., Mooney, H.A., Agard, J., Capistrano, D., Defries, R.S., Diaz, S., Dietz, T., Duraiappah, A.K., Oteng-Yeboah, A., Pereira, H.M., Perrings, C., Reid, W.V., Sarukhan, J., Scholes, R.J., Whyte, A., 2009. Science for managing ecosystem services: Beyond the Millennium Ecosystem Assessment. Proceedings of the National Academy of Sciences 106, 1305-1312.

Chameides, W.L., Lindsay, R.W., Richardson, J., Kiang, C.S., 1988. The role of biogenic hydrocarbons in urban photochemical smog: Atlanta as a case study. Science 241, 1473-1475.

Chaplin-Kramer, R., Sharp, R.P., Weil, C., Bennett, E.M., Pascual, U., Arkema, K.K., Brauman, K.A., Bryant, B.P., Guerry, A.D., Haddad, N.M., Hamann, M., Hamel, P., Johnson, J.A., Mandle, L., Pereira, H.M., Polasky, S., Ruckelshaus, M., Shaw, M.R., Silver, J.M., Vogl, A.L., Daily, G.C., 2019. Global modeling of nature's contributions to people. Science 366, 255-258.

Costanza, R., D'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., Van Den Belt, M., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253-260.

Costanza, R., De Groot, R., Sutton, P., Van Der Ploeg, S., Anderson, S.J., Kubiszewski, I., Farber, S., Turner, R.K., 2014. Changes in the global value of ecosystem services. Global Environmental Change 26, 152-158.

Czajkowski, M., Kretowski, M., 2016. The role of decision tree representation in regression problems – An evolutionary perspective. Applied Soft Computing 48, 458-475.

Diaz, S., Pascual, U., Stenseke, M., Martin-Lopez, B., Watson, R.T., Molnar, Z., Hill, R., Chan, K.M.A., Baste, I.A., Brauman, K.A., Polasky, S., Church, A., Lonsdale, M., Larigauderie, A., Leadley, P.W., van Oudenhoven, A.P.E., van der Plaat, F., Schroter, M., Lavorel, S., Aumeeruddy-Thomas, Y., Bukvareva, E., Davies, K., Demissew, S., Erpul, G., Failler, P., Guerra, C.A., Hewitt, C.L., Keune, H., Lindley, S., Shirayama, Y., 2018. Assessing nature's contributions to people. Science 359, 270-272.

Diaz, S., Settele, J., Brondizio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Ichii, K., Liu, J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnar, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.J., Visseren-Hamakers, I., Willis, K.J., Zayas, C.N., 2019. Pervasive human-driven decline of life on Earth points to the need for transformative change. Science 366, eaax3100.

Diener, E., Ng, W., Harter, J., Arora, R., 2010. Wealth and happiness across the world: material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. Journal of Personality and Social Psychology 99, 52.

Diener, E., Oishi, S., Tay, L., 2018. Advances in subjective well-being research. Nature Human Behaviour 2, 253-260.

Douglas, O., Lennon, M., Scott, M., 2017. Green space benefits for health and well-being: A life-course approach for urban planning, design and management. Cities 66, 53-62.

Eitelberg, D.A., Van Vliet, J., Doelman, J.C., Stehfest, E., Verburg, P.H., 2016. Demand for biodiversity protection and carbon storage as drivers of global land change scenarios. 40, 101-111.

El-Metwally, A., Javed, S., Razzak, H.A., Aldossari, K.K., Aldiab, A., Al-Ghamdi, S.H., Househ, M., Shubair, M.M., Al-Zahrani, J.M., 2018. The factor structure of the general health questionnaire (GHQ12) in Saudi Arabia. BMC Health Services Research 18.

Felipe-Lucia, M.R., Soliveres, S., Penone, C., Manning, P., van der Plas, F., Boch, S., Prati, D., Ammer, C., Schall, P., Gossner, M.M., Bauhus, J., Buscot, F., Blaser, S., Bluthgen, N., de Frutos, A., Ehbrecht, M., Frank, K., Goldmann, K., Hansel, F., Jung, K., Kahl, T., Nauss, T., Oelmann, Y., Pena, R., Polle, A., Renner, S., Schloter, M., Schoning, I., Schrumpf, M., Schulze, E.D., Solly, E., Sorkau, E., Stempfhuber, B., Tschapka, M., Weisser, W.W., Wubet, T., Fischer, M., Allan, E., 2018. Multiple forest attributes underpin the supply of multiple ecosystem services. Nature Communications 9, 4839.

Frey, B.S., Luechinger, S., Stutzer, A., 2010. The Life Satisfaction Approach to Environmental Valuation, in: Rausser, G.C., Smith, V.K., Zilberman, D. (Eds.), Annual Review of Resource Economics, Vol 2, 2010. Annual Reviews, Palo Alto, pp. 139-160.

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 29, 1189-1232.

Gong, P., Liu, H., Zhang, M.N., Li, C.C., Wang, J., Huang, H.B., Clinton, N., Ji, L.Y., Li, W.Y., Bai, Y.Q., Chen, B., Xu, B., Zhu, Z.L., Yuan, C., Suen, H.P., Guo, J., Xu, N., Li, W.J., Zhao, Y.Y., Yang, J., Yu, C.Q., Wang, X., Fu, H.H., Yu, L., Dronova, I., Hui, F.M., Cheng, X., Shi, X.L., Xiao, F.J., Liu, Q.F., Song, L.C., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Science Bulletin 64, 370-373.

Greenwell, B.M., 2017. pdp: an R Package for constructing partial dependence plots. R J. 9, 421.

Hankins, M., 2008. The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. BMC Public Health 8, 355.

Hartig, T., Kahn, P.H., 2016. Living in cities, naturally. Science 352, 938-940.

Jebb, A.T., Tay, L., Diener, E., Oishi, S., 2018. Happiness, income satiation and turning points around the world. Nature Human Behaviour 2, 33-38.

Kopmann, A., Rehdanz, K., 2013. A human well-being approach for assessing the value of natural land areas. Ecological Economics 93, 20-33.

Krekel, C., Kolbe, J., Wuestemann, H., 2016. The greener, the happier? The effect of urban land use on residential well-being. Ecological Economics 121, 117-127.

Lee, A.C.K., Maheswaran, R., 2011. The health benefits of urban green spaces: a review of the evidence. Journal of Public Health 33, 212-222.

Li, C., Managi, S., 2021. Land cover matters to human well-being. Scientific Reports 11.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2, 18-22.

Mackerron, G., Mourato, S., 2009. Life satisfaction and air quality in London. Ecological Economics 68, 1441-1453.

MacKerron, G., Mourato, S., 2013. Happiness is greater in natural environments. Global Environmental Change 23, 992-1000.

Malek, Ž., Verburg, P.H., 2020. Mapping global patterns of land use decision-making. Global Environmental Change 65, 102170.

Managi, S., Kumar, P., 2018. Inclusive wealth report 2018: measuring progress towards sustainability. Routledge.

Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A.M., de Vries, S., Triguero-Mas, M., Brauer, M., Nieuwenhuijsen, M.J., Lupp, G., Richardson, E.A., Astell-Burt, T., Dimitrova, D., Feng, X.Q., Sadeh, M., Standl, M., Heinrich, J., Fuertes, E., 2017. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. Environmental Research 158, 301-317.

Mendoza-Ponce, A., Corona-Núñez, R., Kraxner, F., Leduc, S., Patrizio, P., 2018. Identifying effects of land use cover changes and climate change on terrestrial ecosystems and carbon stocks in Mexico. Global Environmental Change 53, 12-23.

OECD, 2017. Green Growth Indicators 2017.

Quek, K.F., Low, W.Y., Razack, A.H., Loh, C.S., 2001. Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: A Malaysian study. Psychiatry and Clinical Neurosciences 55, 509-513.

Saphores, J.-D., Li, W., 2012. Estimating the value of urban green areas: A hedonic pricing analysis of the single family housing market in Los Angeles, CA. Landscape and Urban Planning 104, 373-387.

Seresinhe, C.I., Preis, T., MacKerron, G., Moat, H.S., 2019. Happiness is Greater in More Scenic Locations. Scientific Reports 9, 11.

Seresinhe, C.I., Preis, T., Moat, H.S., 2015. Quantifying the Impact of Scenic Environments on Health. Scientific Reports 5, 16899.

Tsurumi, T., Imauji, A., Managi, S., 2018. Greenery and Subjective Well-being: Assessing the Monetary Value of Greenery by Type. Ecological Economics 148, 152-169.

Tsurumi, T., Managi, S., 2015. Environmental value of green spaces in Japan: An application of the life satisfaction approach. Ecological Economics 120, 1-12.

Tsurumi, T., Managi, S., 2017. Monetary Valuations of Life Conditions in a Consistent Framework: The Life Satisfaction Approach. Journal of Happiness Studies 18, 1275-1303.

White, M.P., Alcock, I., Wheeler, B.W., Depledge, M.H., 2013. Would You Be Happier Living in a Greener Urban Area? A Fixed-Effects Analysis of Panel Data. Psychological Science 24, 920-928.

White, M.P., Pahl, S., Wheeler, B.W., Depledge, M.H., Fleming, L.E., 2017. Natural environments and subjective wellbeing: Different types of exposure are associated with different aspects of wellbeing. Health & Place 45, 77-84.

Yuan, L., Shin, K., Managi, S., 2018. Subjective Well-being and Environmental Quality: The Impact of Air Pollution and Green Coverage in China. Ecological Economics 153, 124-138.

Appendix:

Table A1: Numbers of Observations in Each Country and Survey Method

Country	Number of Observations	Survey Method	GDP in 2017 (current USD)	Population in 2017
Egypt	1,010	Face-to-Face-Based	2.3537E+11	9.6443E+07
South Africa	1,110	Web-Based	3.4955E+11	5.7000E+07
China	18,931	Web-Based	1.2143E+13	1.3864E+09
India	6,562	Both	2.6522E+12	1.3387E+09
Indonesia	2,363	Both	1.0154E+12	2.6465E+08
Japan	10,098	Web-Based	4.8600E+12	1.2679E+08
Kazakhstan	1,000	Face-to-Face-Based	1.6681E+11	1.8038E+07
Malaysia	1,077	Web-Based	3.1896E+11	3.1105E+07
Mongolia	500	Face-to-Face-Based	1.1426E+10	3.1138E+06
Myanmar	1,073	Face-to-Face-Based	6.6719E+10	5.3383E+07
Philippines	1,672	Web-Based	3.1362E+11	1.0517E+08
Singapore	550	Web-Based	3.3841E+11	5.6123E+06
Sri Lanka	284	Web-Based	8.8020E+10	2.1444E+07
Thailand	1,115	Web-Based	4.5528E+11	6.9210E+07
Turkey	1,954	Web-Based	8.5268E+11	8.1102E+07
Vietnam	1,497	Both	2.2378E+11	9.4597E+07
Czech	1,178	Web-Based	2.1591E+11	1.0594E+07
France	2,130	Web-Based	2.5863E+12	6.6865E+07
Germany	3,165	Web-Based	3.6567E+12	8.2657E+07
Greece	1,358	Web-Based	2.0309E+11	1.0755E+07
Hungary	1,354	Web-Based	1.4151E+11	9.7880E+06
Italy	2,106	Web-Based	1.9570E+12	6.0537E+07
Netherlands	1,371	Web-Based	8.3181E+11	1.7131E+07
Poland	2,218	Web-Based	5.2622E+11	3.7975E+07
Romania	472	Web-Based	2.1170E+11	1.9587E+07

Russia	2,118	Web-Based	1.5786E+12	1.4450E+08
Spain	2,032	Web-Based	1.3093E+12	4.6593E+07
Sweden	1,330	Web-Based	5.4055E+11	1.0058E+07
United Kingdom	2,993	Web-Based	2.6662E+12	6.6059E+07
Canada	1,332	Web-Based	1.6469E+12	3.6543E+07
Mexico	1,669	Web-Based	1.1577E+12	1.2478E+08
United States	10,620	Web-Based	1.9485E+13	3.2499E+08
Australia	2,004	Web-Based	1.3301E+12	2.4602E+07
Brazil	2,255	Web-Based	2.0536E+12	2.0783E+08
Chile	1,174	Web-Based	2.7775E+11	1.8470E+07
Colombia	1,089	Web-Based	3.1179E+11	4.8901E+07
Venezuela *	807	Face-to-Face-Based	1.4384E+11	2.9390E+07
Total	95,571		6.6924E+13	5.1513E+09

The proportion of the world population: 68.58%

The proportion of world GDP: 82.67%

Note: Data on population and GDP are provided by World Bank. World Bank did not provide the GDP of Venezuela in 2017.

Population: https://data.worldbank.org/indicator/sp.pop.totl

GDP: https://data.worldbank.org/indicator/ny.gdp.mktp.cd

GDP of Venezuela: https://countryeconomy.com/gdp/venezuela

Table A1: Descriptive Statistics of Features

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Meantal Health Score	88,730	24.313	6.304	0	21	29	36
DIG	88,730	-0.319	0.718	-0.997	-0.774	-0.138	3.000
Social Class	88,730	2.959	0.822	1	3	3	5
Student Dummy	88,730	0.056	0.230	0	0	0	1
Worker Dummy	88,730	0.532	0.499	0	0	1	1
Company Owner Dummy	88,730	0.021	0.142	0	0	0	1
Government Officer Dummy	88,730	0.032	0.175	0	0	0	1
Self-employed Dummy	88,730	0.076	0.265	0	0	0	1
Professional Job Dummy	88,730	0.034	0.182	0	0	0	1
Housewife Dummy	88,730	0.084	0.278	0	0	0	1
Unemployed Dummy	88,730	0.086	0.281	0	0	0	1
Pleasure	88,730	3.165	0.810	1	3	4	4
Anger	88,730	2.351	0.935	1	2	3	4
Sadness	88,730	2.347	0.954	1	2	3	4
Enjoyment	88,730	3.051	0.830	1	3	4	4
Smile	88,730	3.318	0.777	1	3	4	4
Euthusiastic	88,730	0.422	0.494	0	0	1	1
Critical	88,730	0.168	0.374	0	0	0	1
Dependable	88,730	0.648	0.478	0	0	1	1
Anxious	88,730	0.244	0.430	0	0	0	1
Open to New Experience	88,730	0.506	0.500	0	0	1	1
Reserved	88,730	0.429	0.495	0	0	1	1
Sympathetic	88,730	0.622	0.485	0	0	1	1
Careless	88,730	0.119	0.324	0	0	0	1
Calm	88,730	0.499	0.500	0	0	1	1

Uncreative	88,730	0.177	0.382	0	0	0	1
Urban Center Dummy	88,730	0.688	0.463	0	0	1	1
Urban Area Dummy	88,730	0.145	0.352	0	0	0	1
Rural Area Dummy	88,730	0.167	0.373	0	0	0	1
Income Group	88,730	2.794	0.900	1	2	3	5
Female Dummy	88,730	0.489	0.500	0	0	1	1
Age	88,730	42.852	14.824	18	30	54	99
Self-reported Health	88,730	3.825	0.883	1	3	4	5
Bachelor Dummy	88,730	0.389	0.487	0	0	1	1
Master Dummy	88,730	0.091	0.287	0	0	0	1
PhD Dummy	88,730	0.018	0.134	0	0	0	1
Community Livable	88,730	4.026	0.852	1	4	5	5
Community Attachment	88,730	3.621	1.038	1	3	4	5
Community Safety	88,730	3.017	0.748	0	3	3	4
Children Number	88,730	1.214	1.218	0	0	2	10
Cropland (%)	88,730	13.617	19.073	0.000	0.984	18.058	99.804
Forest (%)	88,730	12.490	19.526	0.000	0.508	15.177	100.000
Grassland (%)	88,730	10.491	14.808	0.000	1.051	14.017	99.316
Shrubland (%)	88,730	1.079	3.884	0.000	0.000	0.359	40.000
Wetland (%)	88,730	0.083	0.268	0.000	0.000	0.047	3.000
Water (%)	88,730	3.383	8.059	0.000	0.016	2.287	50.000
Urban Land (%)	88,730	58.020	33.801	0.000	28.996	88.559	100.000
Bare Land (%)	88,730	0.525	2.165	0.000	0.000	0.164	20.000

Table A2: Descriptions of Features

Aspect	Predictor Name	Descriptions or Questions of Predictors
	Mental Health Score	The output variable
Income	DIG	Because of the economic gap between countries, absolute annual individual income is less effective in classifying the level of human mental health among people from different countries. Therefore, we employ a new variable DIG.
Social Class	Social Class	Which **social class** do you feel that you belong within your country? (5: Upper - 1: Lower)
Job	Student Dummy	Are you a student, now? (1: yes, 0: otherwise)
Job	Worker Dummy	Are you a worker including both full-time and part-time, now? (1: yes, 0: otherwise)
Job	Company Owner Dummy	Are you a company owner, now? (1: yes, 0: otherwise)
Job	Government Officer Dummy	Are you a government officer, now? (1: yes, 0: otherwise)
Job	Self-employed Dummy	Are you self-employed, now? (1: yes, 0: otherwise)
Job	Professional Job Dummy	Are you doing a profesional job, such as professor, lawyer, doctor, now? (1: yes, 0: otherwise)
Job	Housewife Dummy	Are you a housewife or househusband, now? (1: yes, 0: otherwise)
Job	Unemployed Dummy	Are you unemployed, now? (1: yes, 0: otherwise)
Emotion Weekly	Pleasure	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)
Emotion Weekly	Anger	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)

Emotion Weekly	Sadness	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)
Emotion Weekly	Enjoyment	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)
Emotion Weekly	Smile	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)
Personality	Euthusiastic	Do you see yourself as someone who is euthusiatic? (1: yes - 0: otherwise)
Personality	Critical	Do you see yourself as someone who is critical? (1: yes - 0: otherwise)
Personality	Dependable	Do you see yourself as someone who is dependable? (1: yes - 0: otherwise)
Personality	Anxious	Do you see yourself as someone who is anxious? (1: yes - 0: otherwise)
Personality	Open to New Experience	Do you see yourself as someone who opens to new experience? (1: yes - 0: otherwise)
Personality	Reserved	Do you see yourself as someone who is reserved? (1: yes - 0: otherwise)
Personality	Sympathetic	Do you see yourself as someone who is sympathetic? (1: yes - 0: otherwise)
Personality	Careless	Do you see yourself as someone who is disorganized? (1: yes - 0: otherwise)
Personality	Calm	Do you see yourself as someone who is calm? (1: yes - 0: otherwise)
Personality	Uncreative	Do you see yourself as someone who is conventional? (1: yes - 0: otherwise)
Geographical variable	Urban Center Dummy	The data are extracted from ESA. Data source (European Commission): https://ghsl.jrc.ec.europa.eu/datasets.php
Geographical variable	Urban Area Dummy	The data are extracted from ESA. Data source (European Commission): https://ghsl.jrc.ec.europa.eu/datasets.php
Geographical variable	Rural Area Dummy	The data are extracted from ESA. Data source (European Commission): https://ghsl.jrc.ec.europa.eu/datasets.php

Income Group	Income Group	Which **income group** do you feel that you belong within your country? (5: Upper - 1: Lower)
Gender	Female Dummy	Are you female? (1: yes, 0: otherwise)
Age	Age	Please tell us your age.
Physical Health	Self-reported Health	All in all, how would you describe your state of health? (5: very good - 1: very poor)
Education	Bachelor Dummy	Are you with bachelor degree? (1: yes, 0: otherwise)
Education	Master Dummy	Are you with master degree? (1: yes, 0: otherwise)
Education	PhD Dummy	Are you with Ph.D. degree? (1: yes, 0: otherwise)
Living Environmental Aspect	Community Liviability	How livable is your neighborhood? (5: very livable - 1: not livable)
Living Environmental Aspect	Community Attachment	How attached are you to your local community? (4: extremly attached - 1: completely detached)
Living Environmental Aspect	Community Safety	Please tell us about safety of your neighborhood. (4: very safe - 1: very dangerous)
Family Aspect	Children Number	How many children do you have?
Land Cover	Cropland (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Forest (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Grassland (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Shrubland (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Wetland (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Water (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/

Land Cover	Urban Land (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Bare Land (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/

Note: The features that do not mention the data source are from our surveys.