Humanities & Social Sciences Communications

REVIEW

https://doi.org/10.1057/s41599-025-05234-8 OPEN

Check for updates

Assessing environmental determinants of subjective well-being via machine learning approaches: a systematic review

Min Yang[1] & Yuxuan Zou[2,3]✉

Understanding the determinants of subjective well-being (SWB) is crucial for advancing social sciences, particularly in relation to environmental and social factors. Machine learning (ML) techniques have gained popularity in SWB research, yet there is limited synthesis of their current implementation. This systematic review examines the application of ML techniques in assessing determinants of SWB, providing a comprehensive synthesis of 25 studies published up to March 2024. The review highlights the growing use of ML methods, such as random forests, artificial neural networks, and gradient boosting, in understanding the complex, non-linear relationships between environmental factors and SWB. Key environmental determinants identified include service accessibility such as parks, supermarkets, and hospitals, safety feelings, and exposure to air pollution. Additionally, significant social factors, including sociodemographics, emotional predictors, family predictors, and social capital, also influence SWB. The review underscores the value of ML in revealing non-linear relationships and threshold effects, which are particularly useful for policymakers aiming to optimize interventions to enhance public well-being. Analysis of the importance of variables within these models enables policymakers to prioritize interventions that target the most influential factors. However, the review also identifies challenges in the application of ML, particularly in model reporting, improved interpretability techniques, and methodological rigor. These insights provide a foundation for future research aiming to leverage ML to generate more robust and actionable knowledge in well-being studies. To fully harness the potential of ML in SWB research and prevent its misuse, future studies should prioritize model interpretability and focus on translating these insights into actionable policy recommendations.

Introduction

Subjective well-being (SWB) is a vital indicator of overall health and quality of life, reflecting how individuals perceive and experience their lives. As SWB is closely linked to the economic, social, and health circumstances of populations, understanding its determinants is crucial for promoting a thriving society (Layard, 2006). In recent years, there has been a

[1]Hong Kong Polytechnic University, Hong Kong, China. [2]Hong Kong Baptist University, Hong Kong, China. [3]Present address: Hong Kong Polytechnic University, Hong Kong, China. ✉email: Gloriayuxuan@gmail.com

growing interest in understanding the determinants of SWB, with particular attention paid to the environmental factors that shape it. Environmental determinants, which encompass elements of the built, natural, and social environments, such as service accessibility, air quality, noise levels, and social cohesion, play a crucial role in shaping individuals' quality of life (Kim et al., 2020; Pfeiffer and Cloutier, 2016). Understanding these key environmental determinants is essential for developing effective policies and interventions aimed at enhancing population well-being.

SWB is a multifaceted concept encompassing various dimensions of individuals' self-perceived quality of life. It is common to divide SWB into two dimensions: hedonic well-being and eudaimonic well-being (Ryan and Deci, 2001). Hedonic well-being, which focuses on life satisfaction and the balance between positive and negative affects, is commonly measured using the Satisfaction with Life Scale (SWLS) (Diener et al., 1985), the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988), and the Cantril Ladder (Cantril, 1965). Eudaimonic well-being, on the other hand, emphasizes personal growth, purpose, and fulfillment and is often assessed using Ryff's Scales of Psychological Well-Being (PWB) (Ryff, 1989) and the World Health Organization Quality of Life (WHOQOL) (The Whoqol Group, 1998). Several classic works have summarized the literature on factors associated with SWB based on traditional statistical models (Aslam and Corrado, 2011; Diener et al., 2018; Mouratidis, 2021). These reviews highlight key contributors to SWB, including the fulfillment of basic human needs (e.g., shelter and food security), social and emotional needs (e.g., positive feelings from respect), and environmental factors (e.g., access to green space) (MacKerron and Mourato, 2013; Tay and Diener, 2011). Environmental factors, encompassing the natural, built, and social environments have received growing attention in SWB research (Wang and Wang, 2016). Theoretically, this effect has been explored through frameworks such as "person–environment fit"(Kahana et al., 2003) and "spaces of well-being" (Fleuret and Atkinson, 2007). Empirical studies highlight the impact of built, natural, and social environments on SWB, with factors like service accessibility, greenspace, noise annoyance, and social cohesion playing significant roles (Dittmann and Goebel, 2010; Morrison, 2011; Nisbet et al., 2011).

The field of urban studies has witnessed a notable shift from traditional regression-based analyses to machine learning (ML) approaches. This methodological transition is driven by the advantages that ML techniques offer, such as the ability to handle complex, high-dimensional data and reveal non-linear relationships, and interactions between multiple variables that traditional regression methods might overlook (Hindman, 2015). Decision-tree-based models, such as Random Forests (RF) and Gradient Boosting Decision Trees (GBDT), enable the identification of key environmental predictors while capturing non-linear effects and variable interactions without requiring predefined model structures (Tang et al., 2020). Artificial Neural Networks (ANN) further enhance analytical flexibility by learning intricate, multi-layered dependencies, making them well-suited for studying spatial and temporal variations in environmental determinants of SWB (Grebovic et al., 2023). While traditional models, such as non-linear regression and multi-level modeling, can handle certain complexities, they still require predefined assumptions about variable relationships and interaction terms. In contrast, ML methods adaptively uncover patterns from large, multi-source datasets, reducing the risk of model misspecification and enabling the integration of diverse data types. The key distinction between ML and traditional statistical methods lies in their underlying approach: ML is inherently data-driven, whereas traditional methods are hypothesis-driven. ML approaches prioritize pattern recognition and prediction, allowing models to learn relationships directly from data without the need for prior assumptions. This data-driven nature enables ML to uncover hidden patterns from the data, offering new opportunities to uncover insights that might be overlooked by traditional methods.

Given these advantages, the application of ML approaches in studying well-being determinants presents a significant opportunity to advance our understanding of how environmental factors influence SWB (Osawa et al., 2022; Song et al., 2023; Zhang et al., 2023). However, despite its growing adoption in SWB research, ML is not always methodologically justified, raising concerns about its necessity and appropriate implementation. Many studies apply ML despite small sample sizes, which undermines the reliability and generalizability of complex models (Flint et al., 2019; Vabalas et al., 2019). Unlike machine learning applications in big data contexts, most SWB studies rely on structured survey datasets that may not require complex feature selection or non-linear modeling (Zhang et al., 2018). Furthermore, some studies fail to justify their choice of ML over traditional regression-based methods, particularly when variable relationships are well-established or predominantly linear. Without clear methodological justification and proper reporting, ML risks being misapplied as a novelty tool rather than a rigorous analytical approach. To date, discussions on the adequacy of using ML in SWB research remain limited. This gap is significant because, while ML offers distinct methodological advantages, its application also presents challenges that must be critically examined. A systematic review is therefore needed to synthesize the current usage of ML methods in related fields and evaluate whether ML genuinely provides novel insights or simply adds computational complexity without meaningful advantages. Furthermore, comparing findings from both ML and traditional approaches would clarify the added value of ML and provide researchers with a clearer understanding of its strengths, limitations, and appropriate contexts for use.

This study, therefore, aims to synthesize existing research that employs ML approaches to identify environmental determinants of SWB, while also comparing these findings with those obtained using traditional methods. This comparison allows us to discern the unique contributions of ML. Furthermore, the study reviews the use of ML in the selected studies in terms of their justification for adopting ML, interpretation of the results, and data quality. Specifically, this review seeks to (1) provide an overview of the current state of research about ML-based studies on environmental determinants of SWB, (2) assess the strengths and limitations of using ML approaches, and (3) identify key determinants of SWB based on existing literature. Through this, we aim to offer insights into the current application of ML in understanding environmental determinants of SWB and emphasize the importance of avoiding the misuse or inappropriate application of ML approaches in future studies.

Methodology

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). The review focused on identifying and synthesizing studies investigating the environmental determinants of SWB using ML approaches.

Eligibility criteria. The eligibility criteria for selecting studies were defined based on the following inclusion and exclusion parameters:

Inclusion criteria.

1. **Study subject**: Studies that focused on individual-based or regional-based populations.

2. **Study outcome**: Studies that investigated subjective well-being, including perceived mental health, perceived physical health, happiness, life satisfaction, quality of life, etc.
3. **Analytic approach**: Studies employing machine learning methods (e.g., random forest, support vector machine, gradient boosting model).
4. **Article type**: Peer-reviewed research articles.
5. **Time of publication**: Studies published from the inception of the databases up to March 2024.
6. **Language**: Studies published in English.

Exclusion criteria.

1. Studies employing analytic methods rather than machine learning.
2. Studies lacking measures of environmental determinants.
3. Studies treating subjective well-being as an independent variable rather than an outcome.
4. Studies published in a language other than English.
5. Review papers, editorials, commentaries, and non-peer-reviewed articles.

Search strategy and selection process. A systematic search was conducted in four major electronic bibliographic databases: PubMed, Web of Science, PsycINFO, and Scopus. These databases were chosen for their comprehensive coverage of environmental, psychological, and health-related research. Specifically, PubMed and PsycINFO provide extensive access to health-related literature, crucial for understanding how environmental factors impact well-being from health perspectives. Web of Science and Scopus offer interdisciplinary coverage, including environmental, psychological, and social sciences, facilitating a broad understanding of the topic. The search strategy was designed to capture all relevant studies and included combinations of three groups of keywords related to environmental determinants, subjective well-being, and machine learning. Detailed search strategies and keyword combinations used for each database can be found in S1 of the Supplementary Information.

The initial search yielded a total of 449 articles. After removing 116 duplicate records, 333 unique articles were screened based on their titles and abstracts (Fig. 1). Two reviewers (MY and YZ) independently assessed the articles for eligibility based on the predefined criteria. Disagreements were resolved through discussions to reach a consensus. During the title and abstract screening, 230 articles were excluded for reasons such as irrelevant themes (n = 166), use of non-machine learning analytic methods (e.g., Ordinary Least Squares regression; Structural Equation Modeling; and quantitative research designs etc.) (n = 55), lack of measures of environmental or social determinants (n = 4), and studies where SWB was an independent variable (n = 2). This process resulted in 44 articles being selected for full-text screening.

Following full-text screening, another 15 studies were excluded for the following reasons: irrelevant themes (n = 5), use of non-machine learning analytic methods (n = 7) (e.g., Bayesian multi-level ordered logit model, stepwise regression etc.), SWB as an independent variable (n = 1), no full-text access (n = 1), and lack of ML details in the method section (n = 1). Ultimately, 25 articles met all criteria and were included in the final review.

Data extraction and analysis. Data extraction was performed independently by the two primary reviewers (MY and YZ) to ensure consistency and accuracy. Discrepancies in data extraction were resolved through consensus discussions. The two authors followed transparent and systematic approach that included cross-examination and discussions between each step. The extraction process focused on capturing a comprehensive set of details from each study, including the article title and authors, year of publication, sample size (covering both individual and regional-based populations), and sample characteristics such as demographic and geographic details. Additionally, specific machine learning techniques and models employed in the studies were documented, along with the rationale provided by the authors for selecting these ML approaches over traditional methods. The instruments and scales used to assess SWB were carefully recorded and summarized. The extraction process also identified whether the studies reported and intepretated the non-linear relationships between environmental determinants and SWB, highlighting key determinants of SWB based on their relative importance in the ML models.

Given the extensive range of predictive features and SWB outcomes included in the analysis, the studies were deemed to be too heterogeneous for meta-analysis. Consequently, a narrative synthesis was conducted to provide a comprehensive overview of the findings. The studies were initially grouped according to the level of analysis (individual vs. regional) and subsequently according to the type of environmental determinants, based on the differentiation of geographical environments (e.g., built, natural and social environments). The data on the measurement of SWB were coded inductively. All coding decisions were compared between two authors (MY and YZ), with differences being resolved through discussions until consensus was reached.

Author positionality. The biases, histories, and interests of researchers shape their research processes (Bourke, 2014). Acknowledging the positionality of each author is crucial for transparency, as it helps contextualize how research processes and conclusions have emerged from individual perspectives.

The first author, MY, has a background in health geography, specializing in neighborhood health effects and community well-being. This expertise provided a nuanced understanding of how environmental determinants influence SWB, contributing valuable insight throughout the review process. The second author, YZ, has a background in GIS and public health and has published systematic reviews on obesity risks. YZ also specializes in ML analysis and has published journal articles utilizing ML approaches, offering a methodological perspective crucial for analyzing and interpreting complex ML studies. Overall, each author's background and expertise contributed distinct perspectives and skills, ensuring a comprehensive and balanced analysis in the systematic review.

Results

Summary of included studies. Table 1 summarizes the basic information for the included studies. All studies were published between 2018 and 2024, comprising 21 studies based on individual data and four studies utilizing entire regions as the unit of analysis. The majority of these studies were conducted in China (n = 6), followed by the U.S. (n = 4), the U.K. (n = 3), Spain (n = 3), Italy (n = 3), Canada (n = 2), the Netherlands (n = 2), Japan (n = 1), Jordan (n = 1), Iran (n = 1), Poland (n = 1), Austria (n = 1), Slovakia (n = 1), the Czech Republic (n = 1), Germany (n = 1), Ireland (n = 1), Pakistan (n = 1), and the United Arab Emirates (n = 1). The sample sizes in the individual-based studies ranged from 105 to 30,097 participants. For the region-based studies, the units of analysis included counties, boroughs, communities, and social media posts.

Fifteen of the reviewed studies applied machine learning algorithms specifically to address the potential non-linear

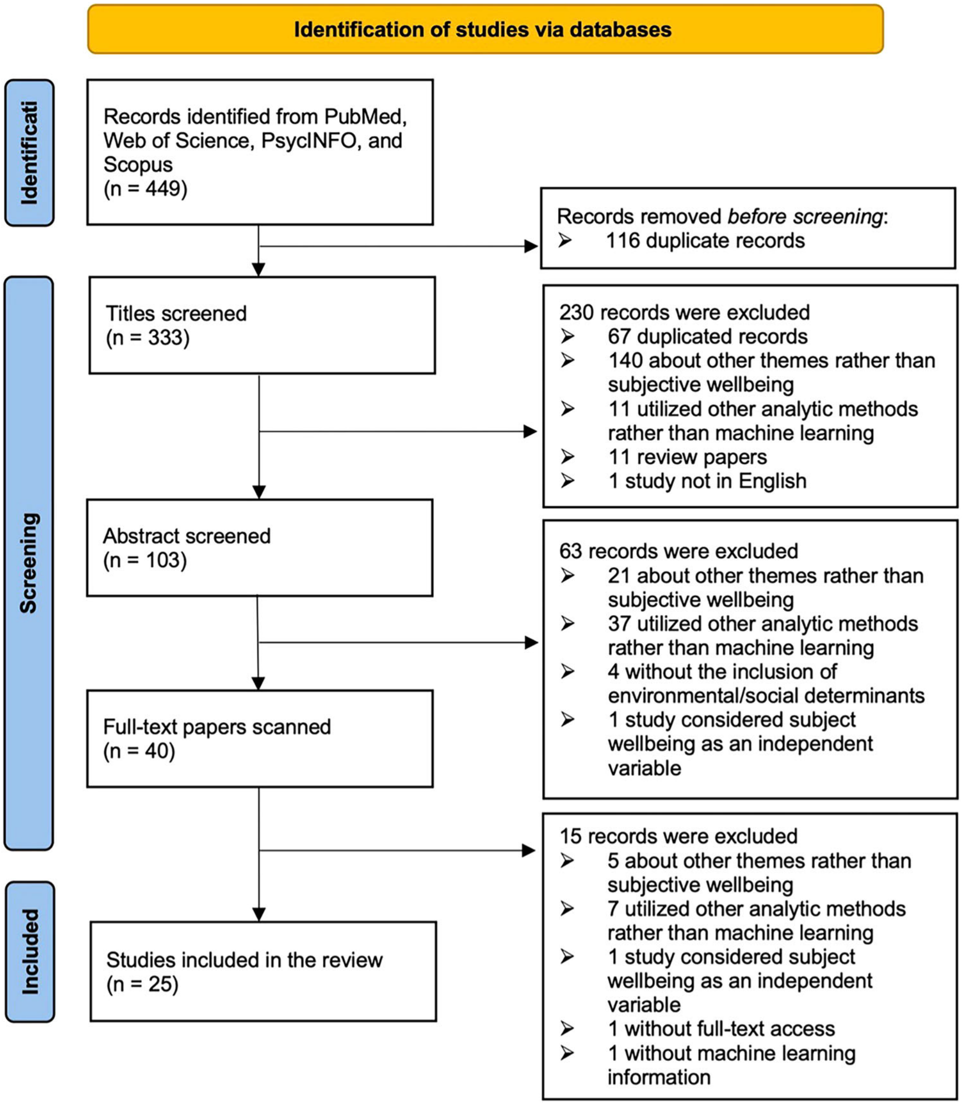

Fig. 1 Study exclusion and inclusion flowchart.

relationships often overlooked by traditional methods. Additionally, 11 studies reported choosing machine learning approaches over traditional statistical models to achieve more accurate estimations. Furthermore, five studies highlighted the ability of ML algorithms to determine the relative importance of various predictors for SWB as a primary reason for favoring ML rather than traditional methods. Notably, two of the reviewed studies did not clearly specify their rationale for using machine learning approaches. Regarding the reporting and interpretation of ML results, only 8 out of 25 studies illustrated and interpreted the identified non-linear relationships using approaches such as partial dependence plots. Interestingly, 6 studies claimed to use ML algorithms to uncover non-linearity issues, yet did not report or interpret any results regarding non-linearity (Cohen and Stutts, 2023; Lin et al., 2021; Osawa et al., 2022; Samani et al., 2020; Schreuder et al., 2016; Wu et al., 2020). As for the reporting of relative importance, 24 out of 25 studies reported such results. The study by Buizza et al. (2022) did not report results of relative importance as was done by other studies. Instead, they primarily focused on identifying the key predictors through regression tree splits without explicitly providing relative importance scores in a numerical or graphical format, thereby limiting the comparability of their findings to those of other reviewed studies.

Methods for measuring SWB. Based on our literature review, we found that 20 out of 25 papers adopted self-report measurements of SWB. Thirteen of these measured eudaimonic well-being, including ten papers on psychological well-being, two on general health, and one on quality of life. Additionally, ten papers measured hedonic well-being, five of which focused on happiness, two on life satisfaction, and three on other measures. Notably, two papers adopted multiple SWB measurements: one measured both perceived general health and successful aging (Mayo et al., 2021), and another measured happiness, life satisfaction, and psychological well-being (Wu et al., 2020).

Apart from papers using self-reported measurements, two studies utilized professional diagnoses of psychotic symptoms (Antonucci et al., 2021) and respiratory disease exacerbations (Samani et al., 2020) as indicators for SWB. Furthermore, three papers employed objectively measured indicators, including public emotion extracted via semantic analysis (He et al., 2024), urban vitality (Ming et al., 2024), and multiple health indicators reflecting length and quality of life (Wei et al., 2022).

Current implementation of machine learning methods. Machine learning approaches differ from traditional linear

Table 1 Overview of reviewed studies.

Reviewed studies (study area, sample size, study level)	Sample characteristics	ML model	Reason for using ML	Measure of SWB	Number of features in ML model	Model performance report	Non-linear relationship	Relative importance	The three most important predictors
Andrejiová et al. (2019) (Slovakia, n = 160, individual)	Respondents aged 23–60	Decision tree	Accurate estimation	Workplace health problems: a feeling of fatigue during working hours, symptoms of spinal pain, headache, feeling cold symptoms, dry nasal mucous membranes.	3	Accuracy = 0.863	N	Y	1. Age 2. Class of work 3. Gender
Antonucci et al. (2021) (Apulia region, Italy, n = 105, individual)	71 health control and 34 psychosis	Support vector machine	Accurate estimation	Psychological well-being: psychotic symptoms according to the Structural Clinical Interview for DSM IV-TR.	6	Balanced accuracy = 72.2, AUC = 0.71, PPV = 50.9%, NPV = 88.0%	N	Y	1. Maternal care 2. Avoidance 3. Maternal overprotection
Buizza et al. (2022) (Brescia, Italy, n = 3754, individual)	Students from Brescia University	Decision tree	To identify the relative contribution of the predictors	1) Psychological well-being measured by the General Health Questionnaire (GHQ). 2) Mental distress measured by the University Stress Scale (USS).	11	GHQ: RMSE = 6.1–11.8, MAE = 1.8–2.4	N	N	1. Physical health 2. Sex 3. Study field
Cohen and Stutts (2023) (U.S., n = 9900, individual)	Adolescents aged 13–17	Artificial neural network	To overcome the non-linearity issue	Psychological well-being: suicidal ideation and attempts.	Suicidal ideation: 2; Suicidal attempts: 3	AUC = 0.68–0.71	N	Y	1. Parental care 2. Family cohesion 3. School engagement
Eder et al. (2021) (Austria, Poland, Spain, Czech Republic, Germany, U.K., Ireland, Italy, and Pakistan, n = 533, individual)	Adult respondents with a mean age of 30.48 years	Least absolute shrinkage and selection operator (LASSO); Extremely Randomized Trees (ERT)	Accurate estimation; to overcome the non-linearity issue	General health: self-reported health status.	26	$R^2 = 0.05$;	N	Y	1. Perceived vulnerability to disease score 2. Exercising 3. Attachment security
Etminani-Ghasrodashti et al. (2021) (U.S., n = 750, individual)	Cancer patients aged over 18	Decision tree; Random forest; and Artificial neural network	Not mentioned	Quality of life: self-reported overall quality of life after cancer treatments.	23	Accuracy = 60–69%, F1-score = 0.69–0.72, AUC = 0.63–0.67	N	Y	1. Age 2. Travel distance to the closest large hospital 3. Perceived accessibility
Helbich et al. (2020) (Netherlands, n = 9435, individual)	Respondents aged 18–65	Artificial neural network; Random forest; and Gradient boosting	To identify the relative contribution of the predictors; to overcome the non-linearity issue	Psychological well-being measured by the Patient Health Questionnaire (PHQ-9).	14	RMSE = 4.48–4.53, MAE = 3.27–3.31, $R^2 = 0.13–0.15$	Y	Y	1. Social cohesion 2. Age 3. Employment status

Table 1 (continued)

Reviewed studies (study area, sample size, study level)	Sample characteristics	ML model	Reason for using ML	Measure of SWB	Number of features in ML model	Model performance report	Non-linear relationship	Relative importance	The three most important predictors
Huang et al. (2023) (China, n = 5736, individual)	Respondents aged 18–90	Gradient boosting	Accurate estimation; to overcome the non-linearity issue; to identify the relative contribution of the predictors	Happiness: single-question measurement of happiness.	18	$R^2 = 0.15$	Y	Y	1. Income 2. Age 3. Cultural facility
Li et al. (2024) (Chengdu, China, n = 515, individual)	Rural older adults over 60 years of age	Random forest	Accurate estimation	Psychological well-being measured by Memorial University of Newfoundland Scale of Happiness.	25	Not mentioned	Y	Y	1. Level of safety near home 2. Supermarket accessibility 3. Health status
Lin et al. (2021) (Abu Dhabi, United Arab Emirates, n = 206, individual)	Occupants of an office building and two academic buildings	Support vector machine; Random forest; and Gradient boosting	Accurate estimation, to overcome the non-linearity issue	Happiness: single-question measurement of happiness.	10	Accuracy = 0.48–0.58, F1-score = 0.36–0.48	N	Y	1. Air quality satisfaction 2. Temperature satisfaction 3. Time lived in UAE
Mayo et al. (2021) (Canada, n = 513, individual)	People with HIV aged 50 years and older	Decision tree	Not mentioned	Successful aging: measured by the RAND-36.	18	Not mentioned	N	Y	1. Loneliness 2. Feeling safe 3. Comorbidity
Morales-Rodríguez et al. (2021) (South-Eastern Spain, n = 337, individual)	University students in South-Eastern Spain	Artificial neural network	Accurate estimation	Psychological well-being measured by the Perceived Stress Scale (PSS).	14	Percentage correct=80%, AUC = 0.748	N	Y	1. Negative self-focus 2. Positive re-evaluation 3. Avoidance
Morgan et al. (2012) (Spain and England, n = 3591, individual)	15-year-old students in 215 schools in Spain and 80 schools in England	Decision tree	To identify the relative contribution of the predictors	Life satisfaction: single-question measurement of life satisfaction.	10	Not mentioned	N	Y	England: 1. Family autonomy and control 2. Family social support 3. School sense of belonging Spain: 1. Family social support 2. School social support 3. Neighborhood sense of belonging
Osawa et al. (2022) (Japan, n = 25,482, individual)	Representative sample of the Japanese population	Random forest; Gradient boosting	To overcome the non-linearity issue	Happiness: single-question measurement of happiness.	30	Not mentioned	N	Y	1. Meaning in life 2. Self-reported poor health 3. Having a spouse
Salameh (2023) (Jordan, n = 240, individual)	Social media users	Artificial neural network	Accurate estimation; to overcome the non-linearity issue	Psychological well-being measured by eight items derived from the Psychological Well-Being Scale (PWB).	4	RMSE = 0.912–1.022	N	Y	1. Self-efficacy 2. Bonding social capital 3. Maintained social capital

Table 1 (continued)

Reviewed studies (study area, sample size, study level)	Sample characteristics	ML model	Reason for using ML	Measure of SWB	Number of features in ML model	Model performance report	Non-linear relationship	Relative importance	The three most important predictors
Samani et al. (2020) (Tehran, Iran, n = 732, individual)	Patients with different types of respiratory disorders	Artificial neural network	To overcome the non-linearity issue	Respiratory disease exacerbations	11	RMSE = 0.017–0.035, R = 0.81–0.91, Precision = 0.95–0.96, Recall = 0.81–0.86, F-score = 0.86–0.91	N	Y	1. Air pollution from CO and PM 2. Amount of exposure to the polluted area 3. Land use
Schreuder et al. (2016) (Netherlands, n = 379, individual)	Patients in four hospitals in the Netherlands	Artificial neural network	To identify the relative contribution of the predictors; to overcome the non-linearity issue	Psychological well-being: Two questions on the patient's mental state.	14	Average error Margin = 0.42	N	Y	1. Safety and security 2. Spatial comfort 3. Autonomy
Song et al. (2023) (Canada, n = 30,097, individual)	Respondents aged 45–85 years	Random forest	Accurate estimation	Psychological well-being measured by self-reported depression diagnosis.	2018	AUC = 0.791 ± 0.016, Balance accuracy = 0.720 ± 0.014	N	Y	1. Personality trait emotional stability 2. Perceived health 3. Social support
Yin and Shao (2021) (China, n = 7837, individual)	Samples from 327 urban communities in China	Gradient boosting	To overcome the non-linearity issue	Happiness: single-question measurement of happiness.	21	Not mentioned	Y	Y	1. Household income 2. Population density 3. Distance to transit
Yu et al. (2023) (Ningbo, China, n = 638, individual)	Respondents aged 16–85 years	Support vector machine; K-Nearest neighbors; Gradient boosting; Random forest; and Extremely random trees regression.	To overcome the non-linearity issue	Travel well-being	26	RMSE = 1.684–3.983	Y	Y	1. Built environment 2. Affective effort 3. Physical effort
Zhang et al. (2023) (Guangzhou, China, n = 1403, individual)	Respondents aged 60 and above	Random forest	Accurate estimation	Psychological well-being measured by the Memorial University of Newfoundland Well-being Scale (MUNSH).	15	MAE = 0.339–1.179, MBE = −0.020–0.159, RMSE = 0.416–1.426, R^2 = 0.581–0.802	N	Y	1. Accessibility to parks 2. Accessibility to supermarkets 3. Accessibility to hospitals
He et al. (2024) (San Francisco, U.S., n = 121,270, regional)	121,270 historical tweets	Random forest	To overcome the non-linearity issue	Public emotion: the semantic orientation of a given text was divided as either positive or negative.	12	MAE = 0.039, MSE = 0.0027, MAPE = 1.092, R^2 = 0.6819	Y	Y	1. Distance to sea 2. Income 3. Noise
Ming et al. (2024) (Chongqing, China, n = 5047, regional)	5047 communities	Random forest	To overcome the non-linearity issue	Urban vitality depicted by the dynamic population density obtained from mobile phone signaling data.	14	R^2 = 0.593–0.640, RMSE = 321.799–770.250, MAE = 207.652–472.661	Y	Y	Urban: 1. Land use mix 2. Floor-area ratio 3. Street sky view index Suburban: 1. Floor-area ratio 2. Land use mix 3. Number of bus stops

Table 1 (continued)

Reviewed studies (study area, sample size, study level)	Sample characteristics	ML model	Reason for using ML	Measure of SWB	Number of features in ML model	Model performance report	Non-linear relationship	Relative importance	The three most important predictors
Wei et al. (2022) (New York State, U.S., n = 620, regional)	62 counties over ten years (2010–2020)	Random forest; Gradient boosting	Accurate estimation; to overcome the non-linearity issue	Length and quality of life: years of potential life lost (YPLL); poor or fair health, poor physical health days, poor mental health days, low birth weight, diabetes, and HIV prevalence.	29	$R^2 = 0.58–0.96$, RMSE = 0.03–0.13	Y	Y	For poor physical health: 1. Teen births; 2. Children in single-parent households; 3. Adult smoking For poor mental health: 1. Year; 2. Free lunch For HIV prevalence: 1. Female population; 2. African American For YPLL: 1. Median household income; 2. Injury deaths
Wu et al. (2020) (Greater London, U.K., n = 96, regional)	32 boroughs over three years (2015–2017)	Support vector machine; Artificial neural network	To overcome the non-linearity issue	Integrated well-being measured by four questions about anxiety, happiness, life satisfaction, and worthiness.	10	MSE = 0.78–0.96	N	Y	2015: 1. Population by nationality; 2. Home fire safety visits; 3. Housing benefit caseload 2016: 1. Dwelling fire fatalities; 2. Dwelling fires; 3. Fire-related fatalities 2017: 1. Private or Light Goods vehicle (PLG) cars; 2. Total PLG; 3. Inland area

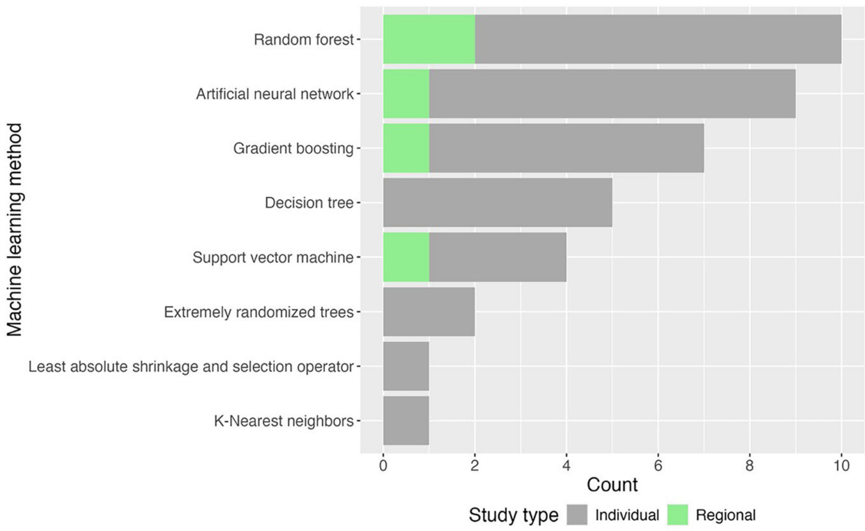

Fig. 2 Machine learning method.

modeling primarily in their data-driven nature, allowing them to identify complex patterns and relationships among variables without relying on predefined assumptions. This flexibility enables ML models to explore relationships within large datasets, uncovering key variables and interactions that might otherwise remain unnoticed.

In the reviewed studies, various ML techniques were employed to explore the determinants of SWB. As shown in Fig. 2, the most frequently employed machine learning method in these studies was the random forest (n = 10), followed by artificial neural networks (n = 9), gradient boosting (n = 7), and decision trees (n = 5). Notably, five studies employed multiple machine learning algorithms and reported the best-performing models based on various accuracy indicators, such as the area under the receiver operating characteristic curve (AUC), root-mean-square error (RMSE), and R^2 (Etminani-Ghasrodashti et al., 2021; Helbich et al., 2020; Lin et al., 2021; Song et al., 2023; Wei et al., 2022).

A notable trend of the reviewed studies was the considerable variation in sample sizes and the number of features included in the ML models. The sample sizes ranged from as few as 96 to over 121,000, reflecting a broad diversity in study design and data availability. Interestingly, more than half of the studies (n = 14) utilized sample sizes smaller than 1000, with three studies including fewer than 200 samples (Andrejiová et al., 2019; Antonucci et al., 2021; Wu et al., 2020). This variation in sample size raises questions about the suitability of ML methods in smaller datasets, as ML models typically require larger samples to avoid overfitting and to ensure reliable results. Regarding the number of features included, most studies (n = 21) inputted more than 10 features into their ML models, reflecting an effort to capture the multidimensional nature of SWB. However, four studies included only 2–6 features (Andrejiová et al., 2019; Antonucci et al., 2021; Cohen and Stutts, 2023; Salameh, 2023), which may limit the ability of the models to fully explore the factors affecting SWB. While these studies reported good model performance, they also face a higher risk of overfitting, especially given the small sample sizes. Conversely, one study initially included 2,018 features in its ML model but ultimately identified 183 as relevant predictors, reporting the top 20 most important factors in their findings (Song et al., 2023). This example illustrates the capacity of ML models to handle high-dimensional data, allowing them to analyze a vast number of potential predictors simultaneously. However, it also underscores the critical role of feature selection in ensuring model interpretability, efficiency, and robustness.

The reviewed studies also exhibited substantial variation in the reporting of model performance. Building on the diversity of ML techniques used, the application of multiple performance metrics in five studies (Etminani-Ghasrodashti et al., 2021; Helbich et al., 2020; Lin et al., 2021; Song et al., 2023; Wei et al., 2022) is a notable strength. By reporting different metrics such as accuracy, AUC, RMSE, and R^2, these studies were able to assess the models from various angles—prediction accuracy, classification performance, and regression error. This approach strengthens the validity of the findings and helps mitigate the limitations of relying on a single metric. For example, AUC is particularly valuable in classification tasks, especially when the data is imbalanced (Wang et al., 2021), while RMSE provides insight into how well the model predicts continuous outcomes (Hodson, 2022). However, there is a noticeable lack of consistency in the metrics used across studies. While accuracy was frequently reported (in 5 studies), it might not always provide a complete picture, especially when dealing with imbalanced datasets or complex models. AUC (in 5 studies) and RMSE (in 8 studies) offer complementary perspectives that can capture different aspects of model performance, but only a few studies utilized these metrics. Moreover, the 5 studies that did not report any performance metrics. Without clear reporting on how well the models performed, it becomes difficult to assess the reliability of the findings.

Key environmental factors influencing SWB. Based on their relative importance, the top three predictors of SWB, regardless of the type, have been selected. Figure 3 demonstrates the selected associations between environmental factors and SWB across the individual-based studies. The reviewed studies cover all aspects of the geographic environment, including the built, natural, and social environments. Associations between the built environment and SWB were more frequently reported. Specifically, the accessibility of services and facilities such as supermarkets, transport, hospitals, and parks positively influenced happiness, psychological well-being, and quality of life (Barykin et al., 2023; Eder et al., 2021; Zhang and Dong, 2023). Additionally, land use mix and population density have been linked to respiratory disease exacerbations and levels of happiness (Samani et al., 2020; Yin and Shao, 2021). Within the natural environment, greater satisfaction with noise level, air quality, and temperature can have a positive influence on happiness (Antonucci et al., 2021), while

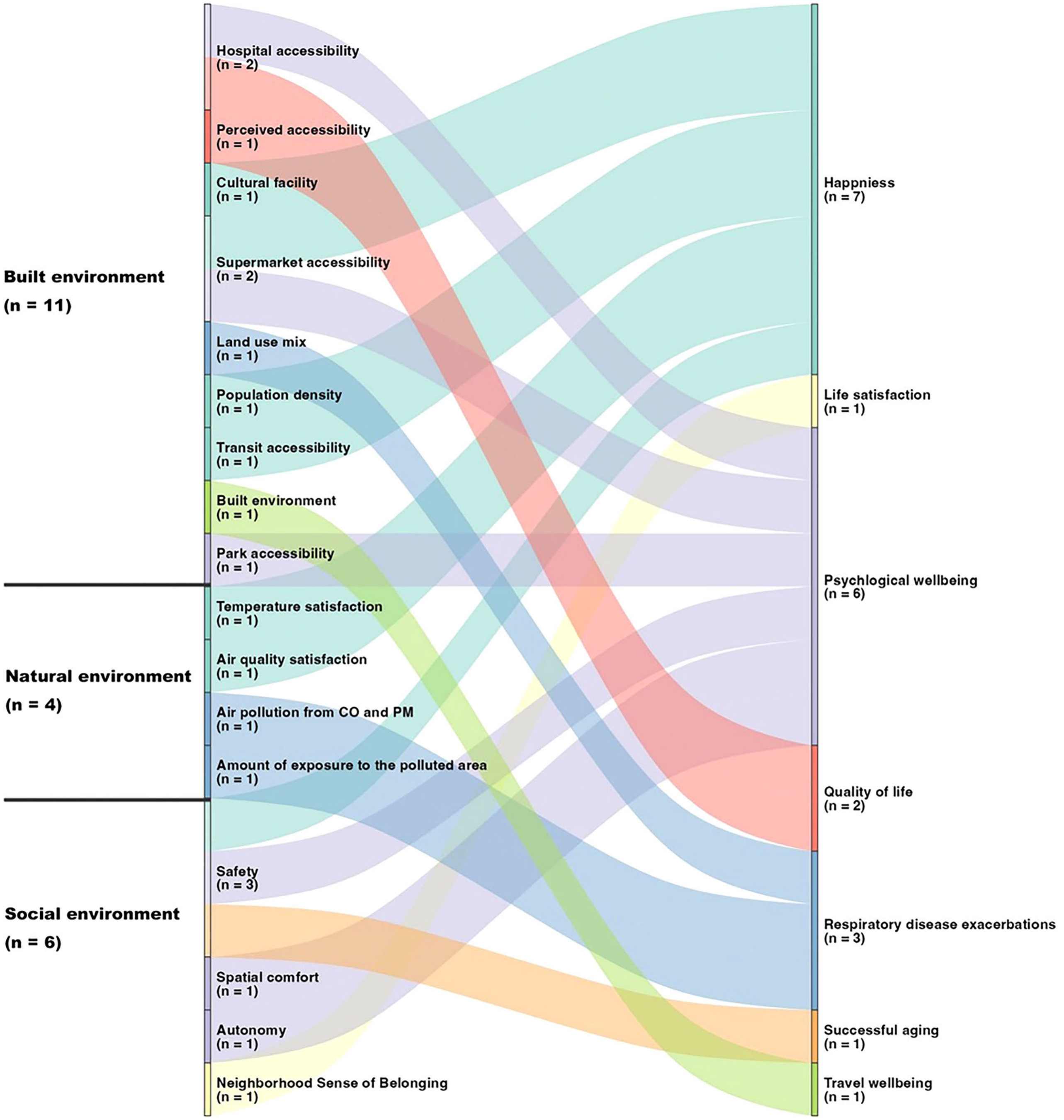

Fig. 3 The identified key environmental factors for SWB.

exposure to air pollution was associated with respiratory disease exacerbations (Samani et al., 2020). Regarding the social environment, feelings of safety were the most commonly reported factor contributing to SWB (Li et al., 2024; Mayo et al., 2021; Schreuder et al., 2016). Additionally, a sense of belonging to one's neighborhood also plays an important role in enhancing life satisfaction (Morgan et al., 2012).

Several studies have provided evidence of non-linear and threshold effects of environmental determinants on SWB (Li et al., 2024; He et al., 2024; Yin and Shao, 2021). For example, Li et al. (2024) and Helbich et al. (2020) highlighted a non-linear relationship between neighborhood safety and SWB. Their findings suggest that when safety conditions reach a moderate level, further improvements in safety lead to substantial increases in well-being. Similarly, studies on access to green space have shown non-linear patterns, with well-being increasing significantly as green space access improves, but reaching a plateau once individuals have sufficient proximity to parks or natural areas (Huang et al., 2023). This suggests that while initial increases in green space accessibility have a substantial impact on SWB, further increases might not contribute significantly beyond a certain threshold.

Additional factors influencing SWB. Among the three main environmental determinants of SWB identified in individual-based studies, 33 social factors were reported (Fig. 4). These factors served as controls features or variables of interests in the reviewed studies. As these factors were found equally and even

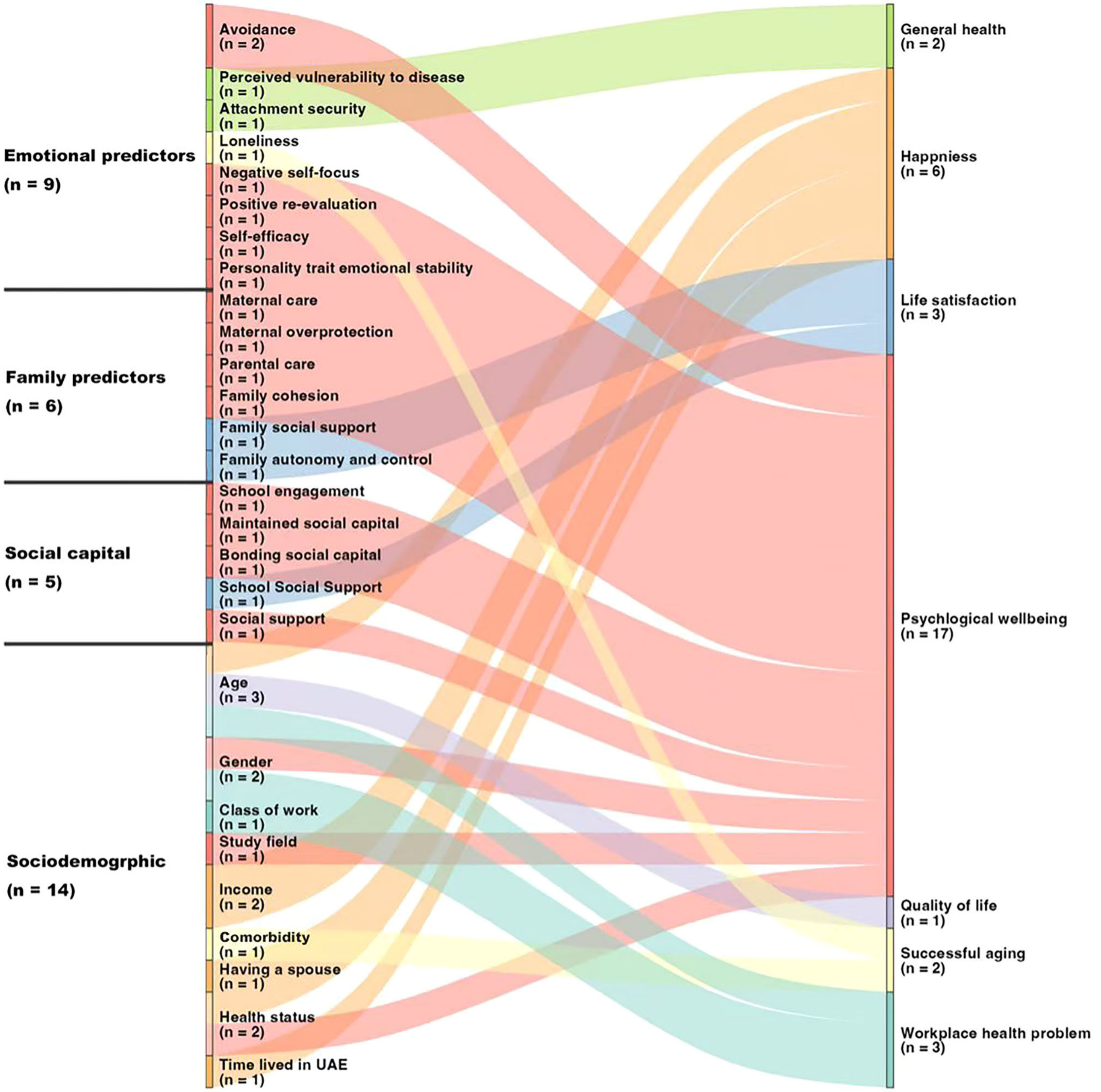

Fig. 4 The identified key social factors for SWB.

more important in explaining individual SWB outcomes (Helbich et al., 2020), we therefore synthesize findings on the social factors for SWB in this review. The social determinants can be categorized into four main groups: sociodemographic factors, emotional predictors, family predictors, and social capital.

For sociodemographic factors, age, gender, and income were commonly reported as key determinants of SWB. Age, in particular, influenced happiness (Huang et al., 2023), quality of life (Etminani-Ghasrodashti et al., 2021), and workplace health problems (Andrejiová et al., 2019). Additionally, income had a positive impact on happiness (He et al., 2024; Huang et al., 2023; Yin and Shao, 2021), as did having a spouse (Osawa et al., 2022).

Emotional factors, including avoidance, loneliness, negative self-focus, positive re-evaluation, and self-efficacy, were also identified as significant predictors of SWB. For example, lower levels of loneliness and a greater sense of safety were beneficial for successful aging, particularly among older individuals living with HIV (Mayo et al., 2021). Furthermore, Salameh (2023) demonstrated that higher levels of self-efficacy, bolstered by social capital, can significantly enhance psychological well-being. On the other hand, perceived stress had a negative impact on SWB, emphasizing the importance of stress management in maintaining overall well-being (Morales-Rodríguez et al., 2021).

Family-related factors, including family autonomy and control, family cohesion, family social support, maternal care, maternal overprotection, and parental care, play a crucial role in predicting life satisfaction and psychological well-being. For instance, inadequate parental care and poor family cohesion were associated with a higher risk of suicidal behavior (Cohen and Stutts, 2023). These family dynamics are vital in shaping individuals' emotional and psychological health (Antonucci et al., 2021; Cohen and Stutts, 2023; Morgan et al., 2012).

Factors related to social capital, such as school engagement, maintained social capital, bonding social capital, and social support, were also identified as important determinants of SWB. Social capital, especially in the form of strong social networks and support systems, significantly contributed to life satisfaction and psychological well-being (Morgan et al., 2012). In addition, school engagement and social support positively influenced life satisfaction and psychological well-being among students and workers (Buizza et al., 2022; Cohen and Stutts, 2023).

Several studies highlight the non-linear relationships and threshold effects of the social determinants of SWB (Helbich et al., 2020; Huang et al., 2023). Huang et al. (2023) analyzed accumulated local effects plots and identified a U-shaped relationship between age and SWB, with older adults reporting higher levels of SWB than middle-aged individuals. They also observed a non-linear association between income and SWB, wherein the positive impact of income plateaued beyond a certain threshold (140,000 CNY/year). Similar patterns were reported by Helbich et al. (2020), who found that middle-aged individuals were at the highest risk of depression, and the protective effect of income against depression diminished for those in the "high" and "very high" income brackets. These findings suggest that while financial stability is crucial for well-being, an excessive emphasis on economic growth does not necessarily lead to further improvements in SWB. Furthermore, the U-shaped relationship between age and SWB highlights the complex variation of SWB across life stages. Middle-aged individuals may experience heightened stress due to work and family responsibilities, whereas older adults, despite potential health concerns, may benefit from greater emotional stability, life satisfaction, and social support, ultimately contributing to higher SWB.

Discussion

This systematic review highlights the significant potential of ML approaches in understanding the determinants of SWB. A comprehensive analysis of 25 studies revealed that ML offers powerful tools to explore environmental determinants of SWB, providing insights that traditional statistical methods might overlook. The ability of ML to handle large, complex datasets and identify non-linear relationships marks a methodological shift in SWB research. However, the application of ML in SWB research is not without challenges, including issues related to model interpretability, the need for large datasets to ensure generalizability, and inconsistencies in performance reporting. Additionally, some studies employ ML without clear methodological justification, potentially prioritizing novelty over rigor. This review summarizes these challenges while illustrating the advantages of ML over traditional methods. We discuss how ML enhances the detection of non-linear relationships and context-dependent interactions, allowing for a deeper understanding of SWB determinants. Furthermore, we emphasize the methodological considerations necessary to improve the robustness, transparency, and applicability of ML-based SWB research, ensuring meaningful advancements in the field.

Summary of the key findings and comparison with traditional research outcomes. This review identified several key determinants of SWB across the studies, categorized into environmental and social factors. Key environmental determinants identified include service accessibility, population density, air quality, and safety. These factors significantly influence both positive and negative emotions and experiences. These findings partially align with those from traditional statistical review studies, which similarly emphasized the importance of green space, neighborhood infrastructure, and safety as primary contributors to SWB (Andalib et al., 2024; Clark et al., 2007; Kodali et al., 2023). While the positive effects of green space have been consistently reported, the associations between other built and social environmental factors and SWB are less consistent (Clark et al., 2007; van Kamp et al., 2003). For example, while some studies suggest that population density exacerbates stress and diminishes well-being (Mouratidis, 2021), others indicate the opposite, highlighting potential benefits such as improved social interactions (Želinský et al., 2021). This inconsistency may stem from the limitations of traditional statistical methods, which, although capable of providing clear and actionable recommendations based on linear relationships, often fail to capture the complex, non-linear, and interactive effects between multiple variables.

In addition to environmental factors, the social determinants, such as sociodemographic factors, emotional predictors, family predictors, and social capital were also frequently reported by the reviewed studies. Age, income, and health status were commonly reported as key factors influencing SWB (Andrejiová et al., 2019; He et al., 2024; Huang et al., 2023; Li et al., 2024). The findings from ML-based analyses align with those obtained from traditional models, where income and health are consistently identified as primary contributors to SWB (Diener et al., 2018; Tay and Diener, 2011).

ML-based studies offer key advantages over traditional statistical methods by uncovering non-linear relationships, identifying intricate variable interactions, and enhancing predictive accuracy. For example, Huang et al. (2023) demonstrated that green space accessibility influences SWB non-linearly, with benefits plateauing beyond a certain threshold, emphasizing the limitations of traditional linear models in capturing saturation points. Furthermore, Helbich et al. (2020) identified pronounced variable interactions for social cohesion, age, employment, and education in predicting depression, illustrating the complex interactions that ML models can uncover. In contrast, traditional statistical methods tend to assume additive effects, potentially leading to oversimplified conclusions about the influence of environmental and social factors on SWB. For instance, Helbich et al. (2020) compared traditional regression models with ML-based approaches and found that regression models failed to capture the non-linear relationship between various physical and social characteristics (e.g., social cohesion and green space) and well-being. Similarly, Wei et al. (2022) demonstrated that linear models often underestimate the interplay between population health outcomes and the predictors, which ML techniques could more accurately model. These findings highlight that ML approaches provide a more sophisticated understanding of the multi-dimensional and context-dependent factors influencing SWB, allowing researchers to move beyond the restrictive assumptions of traditional statistical frameworks.

Furthermore, ML-based studies enhance SWB research by providing detailed rankings of feature importance, offering insights into the most influential factors shaping well-being. Unlike traditional regression models, which determine variable significance through p-values and effect sizes (β-coefficients), ML models rank predictors based on their contribution to overall model performance. This approach enables a hierarchical and context-dependent assessment of influence. For instance, Zhang et al. (2023) applied a RF model to identify key environmental determinants of SWB among the elderly. Their findings revealed that access to parks was the most critical factor in promoting well-being, whereas land use mix—previously shown to have inconsistent associations with SWB in regression-based studies—was not a significant contributor. Similarly, Lin et al. (2021) compared GBDT with ordinal logistic regression and found that ML models assigned greater importance to air quality and temperature satisfaction in predicting individual happiness.

Moreover, Song et al. (2023) utilized ML-based feature importance rankings to identify key predictors of SWB, demonstrating the effectiveness of ML in capturing influential factors across different population subgroups.

Challenges of current ML implementation in SWB research. While the reviewed studies demonstrate the potential of ML in advancing SWB research, several critical gaps must be addressed. A primary concern is the interpretability of ML models. While models like artificial neural networks and gradient boosting offer high predictive accuracy, they often operate as "black boxes," making it difficult to interpret the results and understand the underlying mechanisms driving the predictions (Boelaert and Ollion, 2018). To address these concerns, eXplainable AI (XAI) techniques, such as SHapley Additive Explanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME) and Partial Dependence Plots (PDP), are gaining traction (Lundberg and Lee, 2017; Ribeiro et al., 2016). These methods provide interpretable results by offering insights into how individual variables influence model predictions, thus enhancing transparency. Notably, only eight of the 25 reviewed studies presented the identified non-linear relationships, with even fewer providing detailed interpretations of these findings. Studies such as Huang et al. (2023) and Osawa et al. (2022) have demonstrated the usefulness of these techniques in improving transparency and facilitating the responsible application of ML in SWB research.

Additionally, the reviewed studies highlighted issues related to data quality and availability. Sample sizes varied widely, ranging from 96 to 121,270 participants, with 14 studies using samples of fewer than 1000. Smaller sample sizes can lead to model overfitting, where the model becomes too closely tailored to the training data, capturing noise or random fluctuations rather than the underlying patterns. As a result, the model may perform exceptionally well on the training data but poorly on new, unseen data, leading to unreliable predictions and reduced generalizability. On the other hand, increasing model complexity through large sample sizes and a higher number of variables makes ML models more computationally intensive and difficult to manage, particularly when datasets contain missing values or inconsistent variable definitions. Studies such as Song et al. (2023) emphasize that while ML can efficiently handle large datasets, excessive model complexity may reduce interpretability and limit practical applications in SWB research. In their study, the initial feature set consisted of 2018 variables, but through systematic feature selection, only 183 were ultimately identified as relevant predictors, with the top 20 reported in their findings. While ML algorithms can technically incorporate hundreds or even thousands of features, an excessive number of input variables can introduce noise, increase computational complexity, and lead to overfitting. Therefore, systematic feature selection is essential to retain only the most meaningful predictors and improve model efficiency. For instance, Wei et al. (2022) adopted a Percentile-based Variable Selection approach to filter out less influential variables based on their relative importance, reducing the feature set from 58 to 29. This process not only improved computational efficiency but also enhanced model interpretability by focusing on the most relevant predictors of SWB.

The inconsistent reporting of model performance metrics complicates the evaluation of ML applications in SWB research. Studies used a variety of metrics, including accuracy, AUC, RMSE, and R^2, but there was no standard approach—some studies relied on a single metric, while others reported multiple. This inconsistency hinders direct comparisons between studies in SWB research. Notably, five studies did not report any performance metrics, making it difficult to assess the reliability and generalizability of their findings. Clear and consistent reporting is crucial to ensure replicability and enable meaningful comparisons across studies.

Finally, a significant challenge for current ML studies on SWB is the lack of rigorous justification for selecting ML approaches over traditional statistical models. The review revealed that two out of the 25 studies failed to clearly explain why ML was chosen, and among studies claiming using ML approach to reveal the potential non-linear relationship, 6 studies did not provide results interpreting the non-linear relationships. An uncritical assumption that ML models are inherently superior to traditional regression methods should be approached with caution, as ML-generated associations do not always produce theoretically robust results. For instance, a regional study in London examining environmental influences on mental health identified fire-related variables as key predictors of well-being, despite a lack of clear theoretical justification for this association (Wu et al., 2020). This highlights the importance of integrating domain knowledge into ML research to ensure that findings are both interpretable and theoretically grounded. Researchers must clearly articulate their rationale for selecting ML models, demonstrating the specific advantages gained over traditional methods. Empirical results that substantiate these advantages should be provided, ensuring that the choice of analytical method aligns with the research objectives and enhances the study's overall quality.

Future directions. The findings of this review suggest several avenues for future research. There is a clear need for the development of ML models that strike a balance between predictive accuracy and interpretability. Techniques such as explainable AI could be further explored to enhance the transparency of ML models in well-being research, allowing for more actionable insights. Additionally, longitudinal studies that leverage ML to track the evolution of well-being determinants over time could provide deeper insights into how changes in social and environmental factors influence well-being (Huang et al., 2023). Furthermore, standardizing performance metrics can improve comparability and reproducibility across studies. While some studies reported accuracy, others used metrics like AUC, RMSE, and R^2, leading to inconsistencies in model evaluation. Ensuring the consistent reporting of these metrics will facilitate meaningful cross-study comparisons. Finally, enhancing data quality is critical for improving the robustness and generalizability of ML findings. Utilizing high-quality, multi-source datasets can mitigate biases and reduce overfitting, ensuring that ML applications in SWB research produce reliable and actionable insights.

Strengths and limitations. This systematic review provides a comprehensive synthesis of the existing literature, focusing on the application of ML techniques in understanding the determinants of SWB. By systematically categorizing and analyzing studies that use advanced ML methods, this review highlights the growing importance of these techniques in social science research, offering a clear overview of current trends and methodologies.

However, this review also has limitations. First is the potential for publication bias. The review primarily includes peer-reviewed articles, which may overrepresent studies with positive findings or those that demonstrate the successful application of ML techniques. Additionally, the review is limited by the scope of the databases searched and the inclusion criteria used, which may have excluded relevant studies published in languages other than English or those not indexed in the selected databases. As a result, the grey literatures, such as unpublished manuscripts and government reports, as well as the relevant studies published in languages other than English or those not indexed in the selected databases may have been excluded. To address these limitations, future research should consider incorporating unpublished studies, such as preprints, reports, and government documents, and utilize a broader range of databases to ensure a more comprehensive representation of the field.

Conclusion

In conclusion, this review offers important insights into the application of ML techniques in SWB research, shedding light on both the opportunities and challenges these methods present. While ML provides substantial benefits in analyzing complex datasets and uncovering non-linear relationships, the review also emphasizes significant challenges, including overfitting, data quality, and difficulties in interpretation. By bringing these issues to the forefront, this review paves the way for future research to refine ML approaches, thereby enhancing their utility in advancing our understanding of SWB and informing the development of more effective and equitable policies.

Data availability

No datasets were generated or analysed during the current study.

Received: 24 September 2024; Accepted: 5 June 2025;
Published online: 16 June 2025

References

Andalib E, Temeljotov-Salaj A, Steinert M, Johansen A, Aalto P, Lohne J (2024) The interplay between the built environment, health, and well-being—a scoping review. Urban Sci 8(4):184

Andrejiová M, Piňosová M, Králiková R, Dolník B, Liptai P, Dolníková E (2019) Analysis of the impact of selected physical environmental factors on the health of employees: creating a classification model using a decision tree. Int J Environ Res Public Health 16 (24). https://doi.org/10.3390/ijerph16245080

Antonucci LA, Raio A, Pergola G, Gelao B, Papalino M, Rampino A, Andriola I, Blasi G, Bertolino A (2021) Machine learning-based ability to classify psychosis and early stages of disease through parenting and attachment-related variables is associated with social cognition. BMC Psychol 9:15

Aslam A, Corrado L (2011) The geography of well-being. J Econ Geogr 12(3):627–649

Barykin SE, Sergeev SM, Kapustina IV, Fedotov AA, Matchinov VA, Plaza EDLP, Mottaeva AB, Sharkova AV, Borisova AY, Karmanova AE (2023) Environmental sustainability and digital transformation of socio-economic: quality of life perspective. J Environ Assess Policy Manag 25(1):2350001. https://doi.org/10.1142/S1464333223500011

Boelaert J, Ollion É (2018) The Great Regression- Machine learning, econometrics, and the future of quantitative social sciences. Rev Fr Sociol 59(3):475–506. https://doi.org/10.3917/rfs.593.0475

Bourke B (2014) Positionality: reflecting on the research process. Qual Rep 19(33):1–9

Buizza C, Dagani J, Ferrari C, Cela H, Ghilardi A (2022) Machine learning approaches to identify profiles and predictors of psychosocial discomfort among Italian college students. Salud Ment 45(5):213–226. https://doi.org/10.17711/SM.0185-3325.2022.028

Cantril, H (1965) The pattern of human concerns. Rutgers University Press. https://cir.nii.ac.jp/crid/1130282270933054080

Clark C, Myron R, Stansfeld S, Candy B (2007) A systematic review of the evidence on the effect of the built and physical environment on mental health. J Public Ment health 6(2):14–27

Cohen JR, Stutts M (2023) Interpersonal well-being and suicidal outcomes in a nationally representative study of adolescents: a translational study. Res Child Adolesc Psychopathol 51(9):1327–1341. https://doi.org/10.1007/s10802-023-01068-7

Diener E, Emmons RA, Larsen RJ, Griffin S (1985) The satisfaction with life scale. J Personal Assess 49(1):71–75. https://doi.org/10.1207/s15327752jpa4901_13

Diener E, Oishi S, Tay L (2018) Advances in subjective well-being research. Nat Hum Behav 2(4):253–260. https://doi.org/10.1038/s41562-018-0307-6

Dittmann J, Goebel J (2010) Your house, your car, your education: the socio-economic situation of the neighborhood and its impact on life satisfaction in Germany. Soc Indic Res 96:497–513

Eder SJ, Steyrl D, Stefanczyk MM, Pieniak M, Martínez Molina J, Pešout O, Binter J, Smela P, Scharnowski F, Nicholson AA (2021) Predicting fear and perceived health during the COVID-19 pandemic using machine learning: a cross-national longitudinal study. PLoS ONE 16(3):16. https://doi.org/10.1371/journal.pone.0247997

Etminani-Ghasrodashti R, Kan C, Qaisrani MA, Mogultay O, Zhou HL (2021) Examining the impacts of the built environment on quality of life in cancer patients using machine learning. Sustainability 13(10):5438. https://doi.org/10.3390/su13105438

Fleuret S, Atkinson S (2007) Wellbeing, health and geography: a critical review and research agenda. NZ Geogr 63(2):106–118

Flint C, Cearns M, Opel N, Redlich R, Mehler D, Emden D, Winter N, Leenings R, Eickhoff S, Kircher T, Krug A, Nenadić I, Arolt V, Clark S, Baune B, Jiang X, Dannlowski U, Hahn T (2019) Systematic misestimation of machine learning performance in neuroimaging studies of depression. Neuropsychopharmacology 46:1510–1517. https://doi.org/10.1038/s41386-021-01020-7

Grebovic M, Filipovic L, Katnic I, Vukotic M, Popovic T (2023) Machine learning models for statistical analysis. Int Arab J Inf Technol 20(3A):505–514

He PG, Yu BJ, Ma JX, Luo KQ, Chen ST, Shen ZW (2024) Exploring the non-linear relationship and synergistic effect between urban built environment and public sentiment integrating macro- and micro-level perspective: a case study in San Francisco. Front Psychol 15:1276923. https://doi.org/10.3389/fpsyg.2024.1276923

Helbich M, Hagenauer J, Roberts H (2020) Relative importance of perceived physical and social neighborhood characteristics for depression: a machine learning approach. Soc Psychiatry Psychiatr Epidemiol 55(5):599–610. https://doi.org/10.1007/s00127-019-01808-5

Hindman M (2015) Building better models: prediction, replication, and machine learning in the social sciences. Ann Am Acad Political Soc Sci 659(1):48–62. https://doi.org/10.1177/0002716215570279

Hodson T (2022) Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci Model Dev 7(3):1247–1250. https://doi.org/10.5194/gmd-15-5481-2022

Huang XY, Kang CC, Yin C, Li Y (2023) Urban and individual correlates of subjective well-being in China: an application of gradient boosting decision trees. Front Public Health 11:1090832. https://doi.org/10.3389/fpubh.2023.1090832

Kahana E, Lovegreen L, Kahana B, Kahana M (2003) Person, environment, and person-environment fit as influences on residential satisfaction of elders. Environ Behav 35(3):434–453

Kim ES, Chen Y, Kawachi I, VanderWeele TJ (2020) Perceived neighborhood social cohesion and subsequent health and well-being in older adults: an outcome-wide longitudinal approach. Health Place 66:102420. https://doi.org/10.1016/j.healthplace.2020.102420

Kodali HP, Hitch L, Dunlap AF, Starvaggi M, Wyka KE, Huang TTK (2023) A systematic review on the relationship between the built environment and children's quality of life. BMC Public Health 23(1):2472. https://doi.org/10.1186/s12889-023-17388-8

Layard R (2006) Happiness and public policy: a challenge to the profession. Econ J 116(510):C24–C33. https://doi.org/10.1111/j.1468-0297.2006.01073.x

Li HM, Li MY, Peng PY, Long Y, Ao YB, Bahmani H (2024) Exploring non-linear effects of walking accessibility on well-being in rural older adults of Jintang County: a random forest analysis. Front Public Health 12:1333510. https://doi.org/10.3389/fpubh.2024.1333510

Lin M, Ali A, Andargie MS, Azar E (2021) Multidomain drivers of occupant comfort, productivity, and well-being in buildings: Insights from an exploratory and explanatory analysis. J Manag Eng 37(4):04021020. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000923

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In Proceedings of the 31st international conference on neural information processing systems. 30. https://doi.org/10.48550/arXiv.1705.07874

MacKerron G, Mourato S (2013) Happiness is greater in natural environments. Glob Environ Change 23(5):992–1000

Mayo NE, Brouillette M-J, Nadeau L, Dendukuri N, Harris M, Smaill F, Smith G, Thomas R, Fellows, LK (2021) A longitudinal view of successful aging with HIV: role of resilience and environmental factors. Qual Life Res. https://doi.org/10.1007/s11136-021-02970-7

Ming YJ, Liu Y, Li YP, Yue WZ (2024) Core-periphery disparity in community vitality in Chongqing, China: nonlinear explanation based on mobile phone data and multi-scale factors. Appl Geogr 164:103222. https://doi.org/10.1016/j.apgeog.2024.103222

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morales-Rodríguez FM, Martínez-Ramón JP, Méndez I, Ruiz-Esteban C (2021) Stress, coping, and resilience before and after COVID-19: a predictive model based on artificial intelligence in the university environment. Front Psychol 12:15. https://doi.org/10.3389/fpsyg.2021.647964

Morgan AR, Rivera F, Moreno C, Haglund BJA (2012) Does social capital travel? Influences on the life satisfaction of young people living in England

and Spain. BMC Public Health 12(1):138. https://doi.org/10.1186/1471-2458-12-138
Morrison PS (2011) Local expressions of subjective well-being: the New Zealand experience. Reg Stud 45(8):1039–1058
Mouratidis K (2021) Urban planning and quality of life: a review of pathways linking the built environment to subjective well-being. Cities, 115. https://doi.org/10.1016/j.cities.2021.103229
Nisbet EK, Zelenski JM, Murphy SA (2011) Happiness is in our nature: exploring nature relatedness as a contributor to subjective well-being. J Happiness Stud 12:303–322
Osawa I, Goto T, Tabuchi T, Koga HK, Tsugawa Y (2022) Machine-learning approaches to identify determining factors of happiness during the COVID-19 pandemic: retrospective cohort study. BMJ Open 12(12):e054862. https://doi.org/10.1136/bmjopen-2021-054862
Pfeiffer D, Cloutier S (2016) Planning for happy neighborhoods. J Am Plan Assoc 82(3):267–279. https://doi.org/10.1080/01944363.2016.1166347
Ribeiro MT, Singh S, Guestrin C (2016) Model-agnostic interpretability of machine learning. arXiv preprint arXiv:1606.05386
Ryan RM, Deci EL (2001) On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annu Rev Psychol 52(52):141–166. https://doi.org/10.1146/annurev.psych.52.1.141
Ryff CD (1989) Happiness is everything, or is it? Explorations on the meaning of psychological well-being. J Pers Soc Psychol 57(6):1069–1081. https://doi.org/10.1037/0022-3514.57.6.1069
Salameh AA (2023) Social media-based social capital and psychological well-being: exploring the mediational role of self- esteem. Hell J Psychol 20(2):113–138. https://doi.org/10.26262/hjp.v20i2.8962
Samani ZN, Karimi M, Alesheikh A (2020) Environmental and infrastructural effects on respiratory disease exacerbation: a LBSN and ANN-based spatio-temporal modelling. Environ Monit Assess 192(2):90. https://doi.org/10.1007/s10661-019-7987-x
Schreuder E, Lebesque L, Bottenheft C (2016) Healing environments: what design factors really matter according to patients? An exploratory analysis. Health Environ Res Des J 10(1):87–105. https://doi.org/10.1177/1937586716643951
Song Y, Qian L, Sui J, Greiner R, Li X-M, Greenshaw AJ, Liu YS, Cao B (2023) Prediction of depression onset risk among middle-aged and elderly adults using machine learning and Canadian longitudinal study on aging cohort. J Affect Disord 339:52–57. https://doi.org/10.1016/j.jad.2023.06.031
Tang J, Zheng L, Han C, Yin W, Zhang Y, Zou Y, & Huang H (2020) Statistical and machine-learning methods for clearance time prediction of road incidents: a methodology review. Anal Methods Accid Res 27. https://doi.org/10.1016/j.amar.2020.100123
Tay L, Diener E (2011) Needs and subjective well-being around the world. J Personal Soc Psychol 101(2):354
The Whoqol Group (1998) The World Health Organization quality of life assessment (WHOQOL): development and general psychometric properties. Soc Sci Med 46(12):1569–1585. https://doi.org/10.1016/S0277-9536(98)00009-4
Vabalas A, Gowen E, Poliakoff E, Casson A (2019) Machine learning algorithm validation with a limited sample size. PLoS ONE, 14. https://doi.org/10.1371/journal.pone.0224365
van Kamp I, Leidelmeijer K, Marsman G, de Hollander A (2003) Urban environmental quality and human well-being: towards a conceptual framework and demarcation of concepts; a literature study. Landsc Urban Plan 65(1):5–18. https://doi.org/10.1016/S0169-2046(02)00232-3
Wang F, & Wang D (2016) Place, geographical context and subjective well-being: state of art and future directions. In D Wang, S He (Eds), Mobility, sociability and well-being of urban living. Springer Berlin Heidelberg. pp 189–230 https://doi.org/10.1007/978-3-662-48184-4_10
Wang G, Wong K, Lu J (2021) AUC-based extreme learning machines for supervised and semi-supervised imbalanced classification. IEEE Trans Syst, Man, Cybern Syst 51:7919–7930. https://doi.org/10.1109/TSMC.2020.2982226
Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. J Personal Soc Psychol 54(6):1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063
Wei ZY, Narin AB, Mukherjee S (2022) Multidimensional population health modeling: a data-driven multivariate statistical learning approach. IEEE Access 10:22737–22755. https://doi.org/10.1109/ACCESS.2022.3153482
Wu C, Zheng P, Xu XY, Chen SH, Wang NS, Hu SM (2020) Discovery of the environmental factors affecting urban dwellers' mental health: a data-driven approach. Int J Environ Res Public Health 17(21):8167. https://doi.org/10.3390/ijerph17218167
Yin CY, Shao CF (2021) Revisiting commuting, built environment and happiness: new evidence on a nonlinear relationship. Transp Res Part D - Transp Environ 100:103043. https://doi.org/10.1016/j.trd.2021.103043
Yu HM, Ye XF, Liu LN, Wang T, Yan XC, Chen J, Ran B (2023) Analyzing multi-factor effects on travel well-being, including non-linear relationship and interaction. Transp Plan Technol https://doi.org/10.1080/03081060.2023.2285456
Zhang C, Dong CS (2023) The influence of social support on the mental health of elderly individuals in healthy communities with the framework of mental toughness. Psychol Res Behav Manag 16:2977–2988. https://doi.org/10.2147/PRBM.S413901
Zhang L, Fan Y, Zhang W, Zhang S (2018) Subjective well-being prediction using data mining techniques: evidence from Chinese General Social Survey. Appl Comput Math 7(4):197–202. https://doi.org/10.11648/j.acm.20180704.13
Zhang YW, Luo HZ, Xie JM, Meng XZ, Ye CD (2023) The influence and prediction of built environment on the subjective well-being of the elderly based on Random Forest: evidence from Guangzhou, China. Land 12(10):1940. https://doi.org/10.3390/land12101940
Želinský T, Hudec O, Mojsejová A, Hricová S (2021) The effects of population density on subjective well-being: a case-study of Slovakia. Socio-Econ Plan Sci 78:101061. https://doi.org/10.1016/j.seps.2021.101061

Acknowledgements

The authors gratefully acknowledge support from the Department of Applied Social Sciences of The Hong Kong Polytechnic University.

Author contributions

MY: Conceptualization, Methodology, Formal analysis, Resources, Writing–Original Draft and Editing. YZ: Conceptualization, Methodology, Formal analysis, Visualization, Writing–Original Draft and Editing. These authors jointly approved the final manuscript.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethics approval was not applicable. The study does not involve human participants or their data.

Informed consent

Informed consent was not applicable. The study does not involve human participants or their data.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-05234-8.

Correspondence and requests for materials should be addressed to Yuxuan Zou.