RESEARCH ARTICLE

Using Machine Learning to analyze urban landscapes related to public health and well-being

Ruoxi Wang[1], Dehu Kong[2, *]

[1]College of Arts, Xi'an University of Science and Technology, Xi'an, Shaanxi, China. [2]Xi'an Eurasia University, Xi'an, Shaanxi, China.

Received: January 24, 2025; **accepted:** April 9, 2025.

Urban landscapes including parks, green spaces, streetscapes, and built environments play a pivotal role in shaping and impacting the public health and well-being of urban residents. Given the increasing pollution problems facing cities and the overall living standards in urban areas, it is imperative to comprehend the landscape's effect on health. This research established multimedia environmental monitoring data to evaluate health risks in various urban environments and utilized machine learning approaches to understand the intricate connection between urban landscapes and health. The research examined the various urban landscapes with high, moderate, and low pollution levels. The Enhanced Chimp Optimized Resilient Logistic Regression (ECO-RLR) model was proposed to synthesize multimedia environmental monitoring data to assess public health risk in various urban landscapes. Multimedia environmental monitoring data were collected from multiple sources. The data was preprocessed to handle missing values and outliers and normalize the data for further analysis. The results showed that the proposed method performed well with various evaluation parameters of MAE (5.02 μg/m^3), RMSE (8.42 μg/m^3), and R^2 (0.92). The result demonstrated that the proposed model could effectively estimate health risks in urban environments, adapting to the various landscapes. The reliability of the model in assessing health risks in diverse urban environments was validated by the key indices of the noise pollution index (NPI), thermal comfort index (TCI), and indoor air quality index (IAQI). The study showed that machine learning could be used to analyze urban landscapes that affected people's health and well-being. The outcomes demonstrated that the proposed model adapted to various urban environments and could be used to evaluate the risk of environmental health.

Keywords: urban landscapes; public health; urban environment; indoor air quality index (IAQI); health risk; ECO-RLR.

[*]**Corresponding author:** Dehu Kong, Xi'an Eurasia University, Xi'an 710065, Shaanxi, China. Phone: +86 29 8829 8629. Email: flisscahalane@gmail.com.

Introduction

The urban environment is being accepted not just as visual expression and architecture but also as an instrument with powerful influences on health and education [1]. Lots of challenges arise as cities grow and populations increase including polluted air, noise, reduced size of greens, inadequate open spaces, and lack of recreational facilities. These factors influence the overall well-being of urban citizens and lead to several health complications including respiratory illnesses and stress levels [2]. The idea of creating urban environments for health and the health creating approach have received considerable interest and identical responses from policymakers,

urban designers, and health promoters. The premise of this approach is the notion that urban contexts might be purposefully shaped to boost performance in public health. The use of parks, gardens, and green roofs in the built-up environment has been confirmed to offer numerous advantages [3]. Being close to nature has been known to relay stress, improve mood, and even sharpen the mind's capabilities. Further, green spaces promote physical activity like walking, jogging, and outdoor sports that help in preventing lifestyle diseases like obesity, diabetes, cardiovascular diseases. In addition, other people may be aware of the proper design of the urban environment and respond to such exposure, this availability certainly enhances community interaction, relieves stress, and therefore improves mental well-being [4]. Patterns of the built environment in urban areas also contribute to the prevention and control of environmental health risks. Measures that can be taken include improving air quality, using greenery to reduce the heat island effect in residential areas, and properly treating stormwater in advanced designs to mitigate the adverse health effects of the urban environment. [5]. Further, the direct use of permeable surfaces can be useful to control water runoff and avoid water accumulation and potential water-borne sickness. The integration of public health into the designing of urban landscapes calls for multi-sectorial engagements, in which it implies an integration of urban planning, architecture, public health, and environmental science [6]. Cities may incorporate healthy urban planning by focusing not only on the satisfaction of the population but also on developing better, more sustainable, and resilient environments for future generations. A properly designed urban environment is likely to become an important instrument for the enhancement of human well-being and timely diagnosis of health-related issues in post-industrial societies where the urban population is constantly growing [7]. Urban spatial limitations, difficulties in reconciling social, environmental and economic variables, difficulties in quantifying long-term health consequences and lack of funding for integrated ecological infrastructure are some of the limitations. To establish multimedia environmental monitoring data to evaluate health risks in various urban environments, machine learning (ML) approaches can be used to understand the intricate connection between urban landscapes and health.

Semeraro *et al*. developed a methodological solution with great scalability and implementation cost by combining statistical and computational techniques. The approach supported the identification and measurement of diverse perceptions and their relationship to the presence of landscape features with implementation of public space perception while gathering respondents' sociodemographic data. By using a parameterization of pictures that simultaneously took object identification and semantic segmentation for each object as input, it adopted a discrete choice framework to quantify impressions of the spaces [8]. Suel *et al*. examined a technique depending on semantic segmentation handling of street view images to determine the green view index (GVI) of urban streets. The panoramic view green view index (PVGVI) was suggested to measure the amount of visible street-level greenery, and the outcomes were verified by comparing them with those obtained from traditional assessment and the method. The pseudo-Siamese pyramid network (PSPN) method could identify nearly all the greenery information from the street vision images and determine the PVGVI [9]. Kisvarga *et al*. investigated the connections between urban vegetation and pedestrian activity by measuring the real accessibility of pedestrian to urban greenery and found that it could be accomplished with great accuracy and efficiency using Google Street View (GSV) and semantic segmentation, which allowed GVI to accurately assess how greenery affected human behavior and were significant for investigating the complex link between pedestrian activity and urban greenness from a variety of perspectives, particularly about measurements of greenery [10]. Thompson *et al*. examined if the land use's architectural spatial structure was important for urban heat island

(UHI) using ML techniques. Morphological Spatial Pattern Analysis (MSPA) was used to assess the architectural features of green space. The linear relationships between the UHI level and a group of potentially influencing factors were assessed using the association coefficient. A rapidly urbanizing metropolis demonstrated the important influence of architectural characteristics [11]. Ventriglio *et al*. combined temperatures at night with the percentage of forest cover, quantity of permeable surfaces, and poverty level and revealed that these factors were all highly correlated with UHI. The research investigated possible connections between excessive electricity and heat and past demographic, socioeconomic, and land use aspects using an additional dataset on the spatial pattern of the climate throughout a heat wave [12]. Xia *et al*. proposed a deep learning-based technique for evaluating urban inequality by combining street-level and satellite. Three chosen outputs including income, overcrowding, and environmental deprivation were all quantified in decile classes. The research used Mean Absolute Error (MAE) to compare the performance of proposed multimodal models with similar unimodal [13]. Yudono *et al*. investigated the use of ML technology and large urban geospatial data to complete the landscape evaluation of urban regions and develop a technological method for character assessment of the urban landscape relevant to the block size. With the ring road serving as the barrier, the study discovered that urban landscape characteristics varied. However, each zone displayed a mix of various landscape character proportions. The process of mapping urban landscapes was challenging and complex due to the geographical heterogeneity and spectrum overlaps of urban attributes [14]. Zhang *et al*. implemented support vector machine (SVM) and multilayer perceptron artificial neural network (MLP-ANN) ML classifier techniques, as well as group decisions, tree-based classifiers including gradient boosting (GTB) and random forest (RF) to map urban land use classes using multitemporal and multisensor Landsat data and evaluated their accuracies [15].

This research proposed an enhanced chimp optimized resilient logistic regression (ECO-RLR) model to synthesize multimedia environmental monitoring data to assess public health risk in various urban landscapes by collecting data from various multimedia environmental monitoring sources including air quality, thermal comfort, and noise pollution indices. Feature scaling through the Z-score normalization was used for data pre-processing, and then, ECO-RLR was applied to synthesize the features for health risk prediction in urban environments. The model was further assessed through different parameters to confirm the accuracy when it was used to evaluate risks in different urban settings.

Materials and methods

Data collection

The data collection included gathering comprehensive multimedia environmental data from several urban regions in China, which included key indicators of air quality levels (PM 2.5, CO_2), noise pollution levels, thermal comfort indices, and vegetation coverage in urban parks and green spaces. Data were attained from cities with varying pollution levels including high pollution (region 1), moderate pollution (region 2), and low pollution (region 3) to provide diverse perceptions. At least 1,000 data points from each city were included with a total of 3,000 data points for comprehensive analysis. The dataset encompassed 30 Chinese metropolises including urban areas of 22 provincial capitals. To account for China's vast territorial expanse and the heterogeneity of its landforms, the 30 metropolises were divided into four major geographic regions as the northern region including cities of Harbin, Changchun, Shenyang, Dalian, Beijing, Tianjin, Shijiazhuang, Taiyuan, Jinan, Zhengzhou, and Xi'an, the southern region including Shanghai, Nanjing, Hangzhou, Hefei, Ningbo, Wuhan, Nanchang, Chengdu, Chongqing, Guiyang, Kunming, Nanning, Fuzhou, Guangzhou, and Haikou, the northwest region including Hohhot, Yinchuan, and Urumqi, and the Qinghai-Tibet region including Xining and Lhasa. All data

were processed using Python 3.12 (https://www.python.org/) on Windows 11 with Core i7 processor and 32 GB of RAM.

Data preprocessing

The data preprocessing used Z-score normalisation to convert values into an array of standardised values with a standard deviation of 1 and an average deviation of 0. By normalising the data of noise, air quality, and greenery to enhance distribution across the city environment, it was possible to compare their effects on the health and well-being of people in different urban areas. Z-score normalization for urban construction that involved public health and well-being converted data to mean zero and unit variance, which helped comparing different factors such as air quality and green spaces by adjusting for varying scales. Raw data points that were above or below the population mean were represented by Z-scores, while traditional standardization and normalization techniques calculated the total amount of standard deviations. Ideally, the standard deviation should range between -3 and +3. It normalized data that had been set to a previously specified scale to convert any data with different values to a standard value as follows.

$$z\_score = \frac{w-\mu}{\sigma} \quad (1)$$

where σ was the standard deviation. μ was the average. w was the mean of a particular sample.

Enhanced chimp optimized - resilient logistic regression (ECO-RLR)

The ECO-RLR was an integration of ECO and RLR models with the specific aim of enhancing the environment for well-being in the cities. ECO learned from chimp to search for optimal solutions for some of the potential issues related to the layout of cities including green zones and polluting emissions. RLR performed well in situations with noise and missing values, which added to the practical use of forecasting health outcomes. ECO-RLR improved the efficacy of RLR predictions concerning health outcomes such as respiratory diseases or mental health problems by fine-tuning the optimization technique through chimp optimization. For urban planning, ECO-RLR assisted in the determination of the best course of action in terms of planning choices for public health, resilience, as well as well-being through assessment of factors such as air quality, accessibility to green spaces, and noise pollution.

(1) Resilient logistic regression (RLR)

RLR is an improved statistical model for predicting binary outcomes or states that addresses gaps and noise in the data. In the urban environment, RLR could determine the effects of pollutants on population welfare to provide accurate predictions. Several characteristic features were used to represent the urban landscapes. When the data sample consisted of the urban landscape and characteristic parameters, the LR model assessed the landscape. The covariate variables that influenced the landscape parameters were represented by the vector $W(s) = \{w_1(s), w_2(s), \ldots\ldots, w_n(s)\}$, where n was the number of covariate elements. The conditional probabilities of the non-occurring event ($z_s = 1$) was shown below.

$$O\left(z_s \middle| W(s)\right) = \frac{\exp\left(\beta_0+\beta_1 w_1+\beta_2 w_2+\cdots+\beta_n w_n(s)\right)}{1+\exp\left(\beta_0+\beta_1 w_1+\beta_2 w_2+\cdots+\beta_n w_n(s)\right)} \quad (2)$$

where $\beta_0, \beta_1, \ldots\ldots, \beta_n$ were the coefficients of regression of the covariate factor with $\beta_0 > 0$. The normal and failure stages of the bearing condition were separated at s, whereas $z_s = 0$ replaced the failure state, $z_s = 1$ replaced the normal state. The typical signal characteristics from the condition of the machinery at the moment in question were represented by the covariate variable $W(s) = \{w_1(s), w_2(s), \ldots\ldots, w_n(s)\}$. A nonlinear connection existed among the covariate variable $W(s)$ and the rolling urban landscape z_s. In urban landscapes, $W(s)$ was the covariate factor. The expression for the landscape reliability function ratio was $E\left(s\middle|W(s)\right) = 1 - Q\left(s\middle|W(s)\right)$ function distributing failure accumulation to $Q\left(s\middle|W(s)\right)$ as follows.

$$\frac{q\left(s\middle|W(s)\right)}{1-q\left(s\middle|W(s)\right)} = \exp\left(\beta_0 + \beta_1 w_1(s) + \cdots . + \beta_n w_n(s)\right) \quad (3)$$

The logarithmic form of the highest probability estimation approach was shown in equation (4) and might be utilized to solve the parameters.

$$In[K(A)] = \sum_j \left[z_j BX(s) - In\left(1 + \exp\left(BX(s)\right)\right)\right] \quad (4)$$

The weights of the covariate factors were represented by the intercept $\beta_0, \beta_0, \beta_1, \ldots\ldots, \beta_n$. The value $\beta_j > 0$ indicated the event occurrence likelihood raised as the characteristic parameter w_j raised. $\beta_j = 0$ showed that the unique factor did not affect this model. Given the nonlinearity of the logistics model, $\beta_0, \beta_1, \ldots\ldots, \beta_n$ might be estimated using the highest likelihood estimation technique, where $\hat{\beta}$ represented the highest probability estimate of β, which was the urban landscape function as presented in equation (5).

$$O\left(s\middle|W(s)\right) = \frac{\exp\left(\hat{\beta}_0 + \hat{\beta}_1 w_1(s) + \cdots + \hat{\beta}_n w_n(s)\right)}{1+\exp\left(\hat{\beta}_0 + \hat{\beta}_1 g_1(s) + \hat{\beta}_2 g_2(s) + \cdots + \hat{\beta}_n g_n(s)\right)} \quad (5)$$

The accuracy of the residual life forecast model was weakened by the LR model's inability to adjust to variation and disregard the prior deterioration trend. The ELR was then proposed to consider the urban landscape deterioration pattern while predicting residual life without being affected by variations. The ELR reliability function was shown below.

$$O\left(s\middle|W(s)\right) = \frac{\exp\left(\hat{\beta}_0 + \hat{\beta}_1 g_1(s) + \hat{\beta}_1 g_2 + \cdots + \hat{\beta}_n g_n(s)\right)}{1+\exp\left(\hat{\beta}_0 + \hat{\beta}_1 g_1(s) + \hat{\beta}_2 g_2(s) + \cdots + \hat{\beta}_n g_n(s)\right)} \quad (6)$$

where the function of $g_n(s)$ was $w_n(s)$. The characteristic variables that represented the urban landscape condition could be expressed using the function. The ELR-associated functions were then expressed as follows.

$$g_j(s) = \frac{v(s)}{x(s)} \quad (7)$$

$$v(s) = \alpha_n w_j(s_n) + \alpha_{n+1} w_j(s_{n+1}) + \cdots \alpha_{n+m} w_j(s_{n+m}) + w_j(s) \quad (8)$$

$$\alpha_i = \frac{w_j(s) - w_j(s_i)}{(m+1)w_j(s) - w_j(s_n) - w_j(s_{n+1}) - \cdots - w_j(s_{n+m})} \quad (9)$$

where the difference among the eigenvalues was s. The eigenvalues at s_i related to $\alpha_i (i = n, n+1, \ldots, n+m)$ was the mean of α_i and the associated covariates $X(s)$. w_s was the average value of $v(s)$ during a typical workday. The amplitude and distinctive trends were nearly constant when compared to the initial information as shown in equations (8) and (9), which also showed that the ELR took into consideration the landscape deteriorating tendency and lessened the impact of random fluctuations. According to equation (7), the ELR had a high degree of generality and precision and was not impacted by the production and installation of a particular landscape. The probability function's logarithmic version could be expressed as follows.

$$In[K(A)] = \sum_j \left[z_j BH(s) - In\left(1 + \exp\left(BH(s)\right)\right)\right] \quad (10)$$

The urban landscape remaining life at instant $V(u): s < u < \infty$ was roughly represented as equation (11), assuming the covariate variable $K(s) = F(S - s|S > s)]$ being expected.

$$\hat{K}(s) \approx \frac{1}{\hat{Q}\left(s\middle|V(s)\right)} \int_s^{\infty} \hat{Q}\left(s\middle|V(\tau)\right) d\tau \quad (11)$$

Because it ignored the landscape deterioration trend and only took into consideration the landscape characteristic parameters, the LR model was vulnerable to interference. The residual life forecast result introduced errors in the reliability evaluation. In contrast, the ELR took advantage of the landscape deterioration trend and reduced the impact of unpredictable fluctuations, resulting in more precise predictions.

(2) Enhanced chimp optimization (ECO)

ECO refers to a refined optimization algorithm inspired by the social and foraging behaviour of chimps. It was applied to urban landscapes to improve designs that promoted public health and well-being by optimizing green spaces and resource distribution. CO considered the difficult to accurately determine the global optimum in complicated scenarios. Low search accuracy,

poor global search performance, and a poor capacity to strike a balance between local development and global exploration were all features of the basic CO, which was more significant data that happened to show significant nonlinearity over the whole working range for the parameter identification issues. A visual search mechanism and an elite opposition-based learning method were two optimisation strategies that were included in CO to enhance its performance in tackling those issues. One approach was employed to improve CO exploration performance, and the other was utilized to enhance CO exploitation performance (Figure 1).

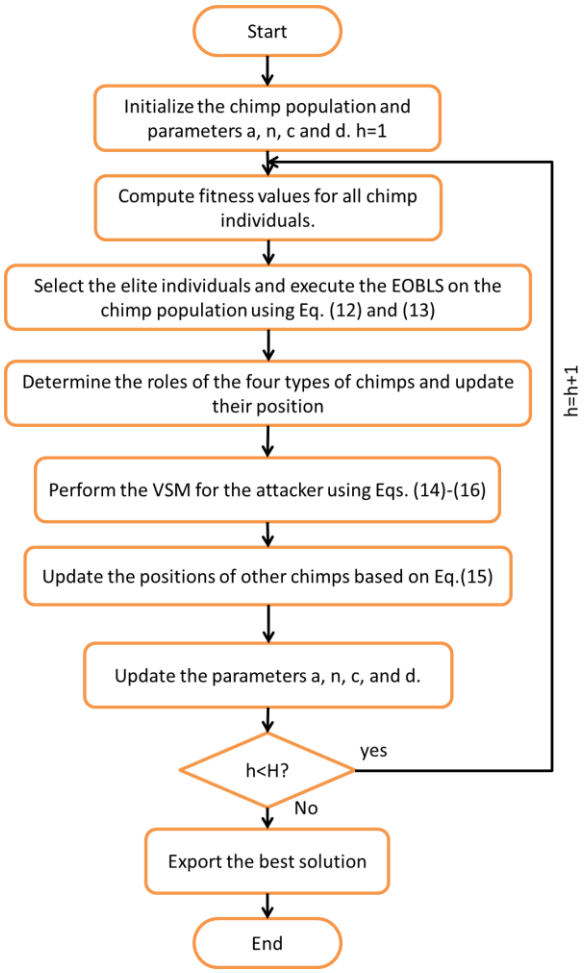

Figure 1. The flow chart of ECO.

Elite opposition-based learning strategy (EOBLS)
Choosing the best responses to pass on to the following generation involves calculating and evaluating both the elite opposing alternatives and the present viable solutions that are equivalent to chimp people at the same time. According to ECO, elite people were those who ranked in the top ten percent of the population in terms of fitness. The elite opposing solution was defined as follows. When $y_j = (y_{j,1}, z_{j,2}, \ldots., z_{j,})\ C$, the solution's elite inversed. $y_j^f = \left(y_{j,1}^f, y_{j,2}^f, \ldots\ldots, y_{j,C}^f\right)$ was calculated as follows.

$$y_j^f = q.(\mu_i + \lambda_i) - y_{j,i}\ ,\ \ j = 1,2, \ldots., M, i = 1,2, \ldots., C \qquad (12)$$

where M was the population size (the total number of chimps). C was the spatial depth. q was a generalized coefficient that indicated a random parameter evenly distributed among $(0, 1)$. μ_i and λ_i were the lowest and highest values of the i-th dimensional parameter for the best solutions, respectively. Notably, the search experience was maintained by using μ_i and λ_i as dynamic bounds that took the place of fixed ones. On the other hand, the elite opposing solutions could emerge from (μ_i, λ_i), and ceased to be viable. In this instance, these responses would be reset using a random generating technique as shown below.

$$y_{j,i}^f = rand(\mu_i, \lambda_i),\ if\ \ y_{j,i}^f < \mu_i\ \ or\ y_{j,i}^f > \lambda_i \qquad (13)$$

This optimisation technique helped CO break from local optimal conditions and enhanced its exploration performance by increasing the population diversity and expanding its exploration space.

Visual search mechanism (VSM)
The top four search agents could gather more detailed information about the ideal solution because this model had a more sophisticated grasp of the prey's position. As a result, there was a good chance that there were other better alternatives close to their locations than the one that was now the best. It provided the attacker

Table 1. IAQI result for impact of public health and well-being in urban landscape.

Pollutant	Low pollution	Moderate pollution	High pollution	Hazardous pollution
PM2.5	20%	35%	65%	90%
CO_2	350 ppm	500 ppm	700 ppm	1,000 ppm
CO	1 ppm	4 ppm	8 ppm	15 ppm
VOC	0.5 mg	1.0 mg	2.0 mg	3.0 mg
Relative humidity	45%	55%	65%	75%

the capacity to visually seek, allowing it to actively look for a better location outside of the passive optimisation that the chimp population induced, which would update its location in time to increase hunting precision and would boost CO exploitation efficiency and convergence accuracy. The process for using the VSM included 3 steps. The step 1 was to determine the maximum optical distance (q_n) among the attack, and each of the three chimp species were shown in equation (12). The step 2 was to create a candidate point $\tilde{y}_b = (\tilde{y}_b^1, \tilde{y}_b^2, \ldots\ldots, \tilde{y}_b^C)$ by actively searching the optical area with a radius of q_n. The step 3 was to choose the better position as the attacker's ultimate location in the current iteration by comparing the fitness metrics of its candidate location and current location, which was represented by equation (14).

$$q_n = \max\{|y_b^i - y_a^i|, |y_b^i - y_d^i|, |y_b^i - y_c^i|\} \quad (14)$$

$$\tilde{y}_b^i = y_b^i + (2.rand - 1).q_n \quad (15)$$

$$y_b = \begin{cases} \tilde{y}_b, & if\ e(\tilde{y}_b) < e(y_b) \\ y_b, & else \end{cases} \quad (16)$$

where $i = 1,2, \ldots, C$. $y_b^i, y_a^i, y_d^i, and\ y_c^i$ was the fitness value. r was a number generated at random in the interval $(0, 1)$. e was the attacker's, obstacle, chaser, and driver, respectively, in their corresponding i -th dimensional spaces.

Validation of ECO-RLR

The proposed model (ECO-RLR) was validated by comparing with the traditional models of genetic algorithm-support vector machine (GA-SVM) [16] and spatiotemporal convolution feature random forest (SCRF) [17]. The parameters of MAE, root mean square error (RMSE), and R^2 were used to determine the effectiveness and accuracy of proposed model.

Results

Indoor air quality index (IAQI)

Urban landscapes often face varying levels of air pollution that directly impact public health and well-being. Pollutants throughout the region included PM 2.5, CO_2, CO, volatile organic compounds (VOCs), and relative humidity played a critical role in determining indoor air quality. The results showed that PM2.5 at low levels (20%) had a minimal effect, but at hazardous levels (90%), could severely affect respiratory health. CO_2 concentrations from 350 ppm to 1,000 ppm indicated escalating risks of carbon accumulation, contributing to climate change. High levels of CO (15 ppm) and VOCs (3.0 mg) led to more severe respiratory and cardiovascular issues. Relative humidity also impacted health with values above 75% linked to discomfort and increased susceptibility to respiratory diseases (Table 1). Therefore, maintaining air quality was essential for enhancing well-being in urban environments.

Thermal comfort index (TCI)

The results showed that the pollution was high in region 1 with the temperature of 35°C, 70% humidity, and 10 km/h wind speed. The TCI comfort level was at 75%, and 45% of the population experienced health issues due to the harsh environmental conditions. Region 2 demonstrated moderate pollution with 28°C,

Table 2. TCI results in the impact of public health and well-being in urban landscape.

Region	Pollution level	Air temperature (C°)	Humidity (%)	Wind speed (Km/h)	Radiant heat (W/m)	TCI comfort level (%)	Population health issues (%)
1	High	35	70	10	500	75	45
2	Moderate	28	60	12	400	60	25
3	Low	25	50	15	350	85	10

60% humidity, and 12 km/h wind speed, which had a comfort level of 60%, and 25% of people reported health problems. Region 3 as a low pollution site showed the temperature of 25°C, 50% humidity, 15 km/h wind speed, which had the higher comfort level at 85%, and only 10% of the population faced health concerns, indicating a positive correlation between environmental quality and public health (Table 2).

Noise pollution index (NPI)

The sound pressure level (SPL) was the highest in region 1 as 25%, indicating significant noise levels, while region 3 experienced a much lower SPL of 5%. The "Duration of exposure" to noise was most continued in region 1 as 30% and decreased in region 2 as 20% and region 3 as 10%. Additionally, the "Frequency spectrum" of noise was higher in region 1 of 20% and lower in region 3 of 10%. “Time of day" showed the highest impact in region 2 with 25% and the lowest in region 1 at 10%. The "Source of noise" had the most substantial effect in region 1 at 40%, while it decreased in other regions. These variables collectively affected the public health outcomes across urban landscapes (Figure 2).

Mean absolute error (MAE)

MAE for the urban landscape health and well-being model measured the average absolute differences between predicted and actual outcomes related to public health factors, indicating the accuracy of the system in forecasting health impacts and well-being needs in urban settings. A lower MAE indicated better prediction performance, leading to more efficient decision-making in urban planning and public health management. The proposed ECO-RLR strategy achieved an MAE value of 5.02 µg/m^3, while the traditional GA-SVM and SCRF approaches yielded higher MAE values of 10.07 µg/m^3 and 6.11 µg/m^3, respectively (Figure 3) (Table 3). The results demonstrated the proposed model’s effectiveness in promoting sustainable and health-conscious urban landscape development.

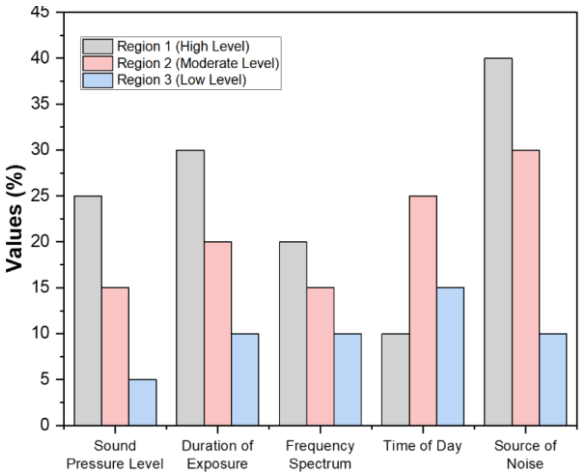

Figure 2. NPI result for impact of public health and well-being in urban landscape.

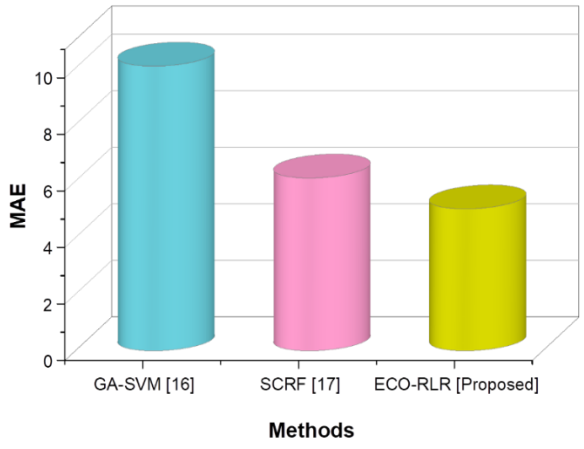

Figure 3. MAE result for public health and well-being in urban landscape.

Root mean square error (RMSE)

The RMSE for the urban landscape health and well-being model quantified the square root of the average squared differences between predicted and actual public health outcomes, which provided a clear measure of prediction accuracy with a low RMSE, indicating better model performance. The proposed ECO-RLR model achieved an RMSE value of 8.42 µg/m^3, whereas the traditional methods of GA-SVM and SCRF resulted in higher RMSE values of 12.1 µg/m^3 and 9.87 µg/m^3, respectively, demonstrating that the proposed model ECO-RLR enhanced effectiveness in public health and well-being (Figure 4) (Table 3).

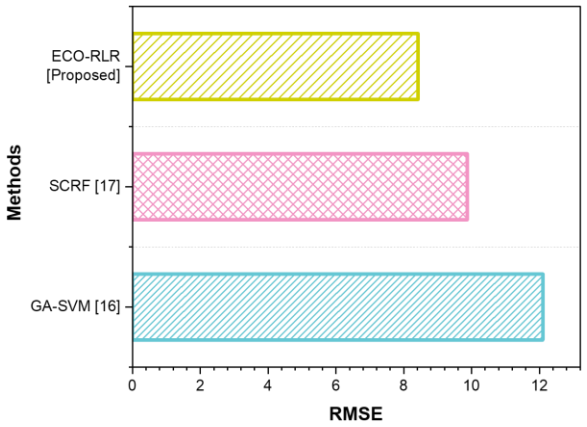

Figure 4. RMSE results for public health and well-being in urban landscape.

R^2 value

The fraction of variability in public health results that could be explained by urban landscape characteristics was shown by the R^2 values for the urban landscapes and public well-being and health system. A higher R^2 suggested a stronger correlation between urban landscape features and overall health and well-being. The proposed ECO-RLR model achieved a value of 0.92, which was higher than that of GA-SVM and SCRF of 0.84 and 0.83, respectively. The results indicated that the proposed model performed noticeably better than current techniques in precisely estimating how urban environments affected people's health and well-being (Table 3).

Table 3. Overall comparison result of MAE, RMSE, R^2.

Methods	MAE (µg/m^3)	RMSE (µg/m^3)	R^2
GA-SVM	10.07	12.1	0.84
SCRF	6.11	9.87	0.83
ECO-RLR	5.02	8.42	0.92

Discussion

Spatial environments are very important in the social determinants of health, which includes availability and access to resources such as clean air and green spaces. The process of urbanization and climate change requires pursuing sustainable design concepts to improve the health and stability of communities. Although the traditional GA-SVM and SCRF models show promises for the optimization of the urban environment for public health and well-being, they both have drawbacks, including that both approaches use large sets of data for training, which may not be always available and provide information about the urban environment. Furthermore, the performance of GA-SVM is affected by over fitting if tuned in inadequately characteristic of diverse urban contexts where multiple factors impact health keys. SCRF performs well in utilizing spatial-temporal features for prediction, but drawbacks could cover data heterogeneity or missing data problems, so it is not flexible for generalizing predictions to other cities or communities. Large data processing is also a requirement for both models, which can become an issue for broad urban health programs. The proposed ECO-RLR method in this study helped to eliminate deficits of the existing GA-SVM and SCRF models through the integration of more progressive optimization procedures and the ability to counter data instability. Thus, the ECO-RLR method decreased over fitting through applied robust regularization and enhanced generalization for different environments. In addition, ECO-RLR had the advantage of scalability, which proved it more useful when applied to large-scale health projects, especially in urban areas and made it the method ideally in real-world urban health analysis. The effect of

urban landscapes on public health and well-being was investigated in this research by integrating multimedia environmental monitoring data with ML methods. Data was collected from multiple sources that related to environmental factors including air quality, thermal comfort, and noise pollution. The datasets were preprocessed to address missing values and outliers and to be normalized for further analysis. The investigation proposed and employed the ECO-RLR model designed to synthesize this multimedia data and evaluate public health risks across various urban landscapes by considering different pollution levels. The proposed ECO-RLR model demonstrated strong performance in estimating health risks comparing to existing traditional models with evaluation parameters of MAE, RMSE, and R^2, confirming its reliability. Additionally, the model incorporated crucial health risk indices including IAQI, TCI, and NPI to further validate its effectiveness in assessing health risks in different urban environments. Urban space limits, the difficulty of harmonizing social, environmental and economic variables, the difficulty of quantifying long-term health consequences, and the lack of financing for comprehensive ecological infrastructure are some of the limitations. Further research and development could be focused on whether using real-time environmental data or incorporating deep learning models to improve health risk assessment. However, to enhance and strengthen the current and similar urban health assessments, more complex and diverse urban contexts should be included, and socioeconomic characteristics need to be examined.

References

1. Fadhil M, Hamoodi MN, Razzak A. 2023. Mitigating urban heat island effects in urban environments: Strategies and tools. IOP conference series. 1129(1):012025.
2. Kay CAM, Rohnke AT, Sander HA, Stankowich T, Fidino M, Murray MH, *et al*. 2021. Barriers to building wildlife-inclusive cities: Insights from the deliberations of urban ecologists, urban planners and landscape designers. People Nat. 4(1):62–70.
3. Lin J, Qiu S, Tan X, Zhuang Y. 2023. Measuring the relationship between morphological spatial pattern of green space and urban heat island using machine learning methods. Build Environ. 228(2):109910.
4. Lu Y, Xu S, Liu S, Wu J. 2022. An approach to urban landscape character assessment: Linking urban big data and machine learning. Sustain Cities Soc. 83(6):103983.
5. Ouma YO, Keitsile A, Nkwae B, Odirile P, Moalafhi D, Qi J. 2023. Urban land-use classification using machine learning classifiers: Comparative evaluation and post-classification multi-feature fusion approach. Eur J Remote Sens. 56(1):2173659.
6. Ramírez T, Hurtubia R, Lobel H, Rossetti T. 2021. Measuring heterogeneous perception of urban space with massive data and machine learning: An application to safety. Landsc Urban Plan. 208(5):104002.
7. Saverino KC, Routman E, Lookingbill TR, Eanes AM, Hoffman JS, Bao R. 2021. Thermal inequity in Richmond, VA: The effect of an unjust evolution of the urban landscape on urban heat islands. Sustainability. 13(3):1511.
8. Semeraro T, Scarano A, Buccolieri R, Santino A, Aarrevaara E. 2021. Planning of urban green spaces: An ecological perspective on human benefits. Land. 10(2):105.
9. Suel E, Bhatt S, Brauer M, Flaxman S, Ezzati M. 2021. Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. Remote Sens. Environ. 257(2):112339.
10. Kisvarga S, Horotán K, Wani MA, Orlóci L. 2023. Plant responses to global climate change and urbanization: Implications for sustainable urban landscapes. Horticulturae. 9(9):1051.
11. Thompson S, Rahmat H, Marshall N, Steinmetz-Weiss C, Bishop K, Corkery L, *et al*. 2023. Merging smart and healthy cities to support community wellbeing and social connection. Encyclopedia. 3(3):1067–84.
12. Ventriglio A, Torales J, Castaldelli-Maia JM, De Berardis D, Bhugra D. 2020. Urbanization and emerging mental health issues. CNS Spectrums. 26(1):1–8.
13. Xia Y, Yabuki N, Fukuda T. 2021. Development of a system for assessing the quality of urban street-level greenery using street view images and deep learning. Urban Forestry and Urban Greening. 7(2):126995.
14. Yudono A, Afrianto F, Santosa H. 2024. Mapping nature's canopy: Analyzing Google Street View's big data for green view index identification. Evergreen. 11(2):1190–1200.
15. Zhang J, Hu A. 2022. Analyzing green view index and green view index best path using Google Street View and deep learning. J Comput Des Eng. 3(1):45-65.
16. Zhang P, Ma W, Wen F, Liu L, Yang L, Song J, *et al*. 2021. Estimating PM2.5 concentration using the machine learning GA-SVM method to improve the land use regression model in Shaanxi, China. Ecotoxicol Environ Saf. 225(4):112772.
17. Chen W, Zhang N, Bai X, Cao X. 2024. Research on the estimation of air pollution models with machine learning in urban sustainable development based on remote sensing. Sustainability. 16(24):10949.