
There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

https://eprints.gla.ac.uk/260536/

Deposited on: 17 January 2022
Predicting the effect of street environment on residents’ mood states in large urban areas using machine learning and street view images

Abstract

Background: Researchers have demonstrated that the built environment is associated with mental health outcomes. However, evidence concerning the effects of street environments on mood in fast-growing Asian cities is scarce. Traditional questionnaires and interview methods are labor intensive and time consuming and pose challenges for accurately and efficiently evaluating the impact of urban-scale street environments on mood.

Objective: This study aims to use street view images and machine learning methods to model the impact of street environments on mood states in a large urban area in Guangzhou, China, and to assess the effect of different street view elements on mood.

Methods: A total of 199,754 street view images of Guangzhou were captured from Tencent Street View, and street elements were extracted by pyramid scene parsing network. Data on six mood state indicators (motivated, happy, positive-social emotion, focused, relaxed, and depressed) were collected from 1590 participants via an online platform called Assessing the Effects of Street Views on Mood. A machine learning approach was proposed to predict the effects of street environment on mood in large urban areas in Guangzhou. A series of statistical analyses including stepwise regression, ridge regression, and lasso regression were conducted to assess the effects of street view elements on mood.

Results: Streets in urban fringe areas were more likely to produce motivated, happy, relaxed, and focused feelings in residents than those in city center areas. Conversely, areas in the city center, a
high-density built environment, were more likely to produce depressive feelings. Street view elements have different effects on the six mood states. “Road” is a robust indicator positively correlated with the “motivated” indicator and negatively correlated with the “depressed” indicator. “Sky” is negatively associated with “positive-social emotion” and “depressed” but positively associated with “motivated”. “Building” is a negative predictor for the “focused” and “happy” indicator but is positively related to the “depressed” indicator, while “vegetation” and “terrain” are the variables most robustly and positively correlated with all positive moods.

Conclusion: Our findings can help urban designers identify crucial areas of the city for optimization, and they have practical implications for urban planners seeking to build urban environments that foster better mental health.

Keywords:
Street view, Mood state, Mental health, Urban residents
1. **Introduction**

Urbanization has become one of the significant global phenomena. The World Urbanization Prospects report revealed that over two-thirds of the world population is expected to live in cities by 2050 (United Nations, 2018). Although the urbanization process improves human living conditions in many dimensions, it has a considerable impact on urban residents’ mental health in terms of mood states such as anxiety and depression. According to the latest nationwide survey, depression and anxiety are on the rise in China, and approximately 16.6% of Chinese adults have experienced mental illness in their lives (Huang et al., 2019). The rapid pace of urbanization and modern city life have given rise to poor mental health for residents in Chinese cities (Chen et al., 2014; Chen and Chen, 2015; Li and Liu, 2018).

Recent research has suggested that complex environments, including high-density buildings and infrastructure in large cities, are more likely to lead people to experience negative moods (Evans, 2003; Gifford, 2007; Morris and Guerra, 2015; Zhu and Fan, 2018). Streets have increasingly attracted research and attention because they are present in all parts of a city, constituting the most common daily landscape for urban residents. China has experienced the most rapid urbanization in the world. During the past decades, numerous investments have been made in urban street network construction to promote the rapid expansion of cities (Zhang and Zhao, 2009). In most Chinese cities, urban streets play an important role in residents’ mood states, as streets are among the most frequently used public spaces, and people spend most of their time in the street environment as they pursue daily outdoor activities (e.g., social interaction, walking and shopping) (Flock and Breitung, 2016; Newton et al., 2010; Sun et al., 2020). Therefore, exploring the complex associations between the street environment and residents’ mood states in large cities in large Chinese cities is important for building urban environments conducive to psychological health.
2. Literature review

Mood states, including positive and negative aspects, can be influenced by various physical settings. Most studies have demonstrated that urban environments, such as parks, gardens, rivers, and neighborhood characteristics (e.g., recreational resources, walkability and greenness), are directly or indirectly linked to mood states in urban populations through their impact on physical activity, social interaction, and exposure to green space; thus, these characteristics impact whether urban environments reduce stress, enhance positive emotions, and allow residents to relax (Hartig et al., 2003; Houlden et al., 2018; Li and Sullivan, 2016; Nordh et al., 2011; Nordh and Østby, 2013; Subiza-Pérez et al., 2020; Wood et al., 2017). Other studies have likewise revealed that the street environment plays an important role in residents’ mood states (Hidalgo, 2021; Marselle et al., 2020; Taylor et al., 2015). Nevertheless, most researches on streets’ impact on mood have focused on vegetation. For instance, streets with a high density of trees and plants may help to alleviate depressive symptoms (Marselle et al., 2020) and restore attention capacity (Zhao et al., 2020). The relationship between other street view characteristics (e.g., building, sky, car, sidewalk) and mood has yet to be investigated. In addition, the impact of street environment on other mood-related aspects such as happiness, motivation and relaxation have seldom been considered in existing research. Thus, more concrete and cohesive studies on the psychological effects of street environment are needed.

Regarding the methods adopted to measure the impact of environments on mood states, most previous epidemiological studies have been carried out using surveys, including questionnaires, interviews, and site observations (Ma et al., 2020; Van den Berg et al., 2014; Wang et al., 2016; White et al., 2010). For instance, the multi-dimensional mood questionnaire and perceived restoration scale have been used in wide variety of studies to measure the
correlations between mood experiences and different environment types (Jiang et al., 2021; Wilkie and Clements, 2018). With the development of electronic technology, many recent studies have used wearable sensor devices to measure people’s stress, moods and preferences through physiological response signals elicited by the visual perception of environmental stimuli, including pulse rate, eye movements, electrodermal activity, heart rate variability, electrocardiograms, electroencephalograms and functional magnetic resonance imaging (Li et al., 2020; Ojha et al., 2019; Qin et al., 2013; Tost et al., 2019; Yin et al., 2020). However, most of the aforementioned studies have been conducted either in real-world, on-site situations or in simulated laboratory environments, which are largely limited and nonrepresentative and constrain the scale of the study area and small sample size. Moreover, the main drawback of these measurement approaches is that they are labor intensive, expensive and time consuming.

With recent advances in sensing technologies, some map services companies (e.g., Google, Tencent and Baidu) have collected a wealth of panoramic street view images at eye level linked to GPS data (Helbich et al., 2019; Rzotkiewicz et al., 2018). These street view images on are freely available on the Internet, and they have become a new source of data for exploring the association between the urban environment and human health. The availability of these datasets with geotagged images has enabled large-scale studies to capture the characteristics and qualities of urban built environments at the microlevel, for example, by identifying street composition variation (Tang and Long, 2019), estimating the sky view factor (Zeng et al., 2018), assessing the shade provision of street trees (Li and Ratti, 2018), measuring access to street greenery and physical activity (Lu, 2019; Lu et al., 2018; Ye et al., 2019), mapping urban perception (Zhang et al., 2018), visually evaluating neighborhood walkability (Zhou et al., 2019), and understanding social sensing (F. Zhang et al., 2019). Although some recent works have shifted toward
evaluating the association between the street environments in large urban areas and residents’ mental health by using street-level image data (Bader et al., 2017, 2015; Li et al., 2015), mainstream literature has largely focused on evidence from cities in North America and Europe, and little research has been carried out in Asian cities (Rzotkiewicz et al., 2018). Several other recent studies were conducted in China but were restricted to a specific elderly population at the neighborhood level (Helbich et al., 2019; Wang et al., 2019), therefore limiting their generalizability with respect to the street environment’s effects on mood in a larger urban environment.

Moreover, many crowdsourcing platforms based on web games have attempted to collect users’ ratings of urban aesthetics, “scenicness,” and perceptions of random geotagged images; these platforms include UrbanGems, Scenic-Or-Not, and Place Pulse(Dubey et al., 2016a; Quercia et al., 2014; Seresinhe et al., 2015). However, there is a paucity of crowdsourcing data characterizing the street environment effects on mood. Emerging machine learning techniques support automatic, effective extraction of street view image features with high-level information and model training to predict image labels at an urban or even global scale (Dubey et al., 2016a; Zhang et al., 2018). Therefore, following this trend, there is high potential that these new data and machine learning algorithms can be employed to measure the impact of street environment on mood in a larger urban area.

In this study, we use street view images and machine learning methods to assess the impact of the urban-scale street environment on residents’ mood states in Guangzhou, China. We investigate the following two research questions: 1) Do different street environments lead to diverging mood states in Guangzhou? 2) What are the key street view elements that impact residents' mood states? We hypothesize that streets situated on high-density and highly urbanized
area relate to more negative moods and that the natural characteristics of the street positively affect residents’ moods. This study extended previous work by studying the association between the street environment and mood at an urban scale using machine learning and streetscape images. In addition, it also has practical implications for urban planners seeking to build urban environments conducive to improving mental health.

3. Method

3.1. Study area

The study was conducted in Guangzhou, China, from September to December 2020. Guangzhou is one of the largest cities of the Pearl River Delta metropolitan area, and 11 districts are under its jurisdiction. In 2019, the city covered a total area of 7,434 km², with an urbanization rate of 86.46% and a permanent resident population of 15.31 million (Guangzhou Statistics Bureau, 2020a, 2020b). According to the rate of mood disorders in China in 2017 (Que et al., 2019), 700 thousand citizens in Guangzhou may suffer from mental illnesses. The Tianhe, Haizhu, Yuexiu, and Liwan districts are Guangzhou’s central districts (Fig. 1). These four districts cover the most densely populated areas of Guangzhou and reflect a representative urban landscape, with an area of 279.63 km² and approximately 6 million inhabitants (Guangzhou Statistics Bureau, 2020b). Thus, they constitute an appropriate study area for exploring the association between the urban environment and residents’ moods.
3.2. Research framework

Fig. 2 illustrates this study’s three main stages. First, street view images of the study area were collected via Tencent Street View through an Application Programming Interface (API) and OpenStreetMap (OSM). Sample images were then selected manually and randomly for use in assessing the perceptual mood states of street environments. Data on six mood state indicators were collected through an online platform from local participants, who scored two-image comparisons. Second, the mood state score of each image was quantified using the strength of schedule method, and, for training the machine learning models, the binary label method was used for image classification. Street-view element extraction was then processed through pyramid scene parsing network (PSPNet) to obtain objective and accurate image segmentation. Third, five machine learning algorithms were applied to test the models’ accuracy and generalizability, and
the algorithm with the most stable performance was ultimately chosen as the pretrained model to predict the effects of urban streets on mood across the study area. Additionally, three statistical analysis methods were used to identify the driving factors underlying streets’ impact on moods.

3.3. Data collection

Street view images of Guangzhou between 2018 and 2020 were captured via Tencent Street View (https://map.qq.com/) through an API. The geographical coordinates were sourced from OSM (https://www.openstreetmap.org/). A total of 199,754 images were obtained at an interval of 20 meters across the whole study area. For each image, the meta-data, including the camera angles of 0, 90, 180, and 270, and the geographical coordinates were collected. Parameters of these images were as follows: size, 512 x 512; horizontal field of view, 90 degrees.

We built an online platform called Assessing the Effects of Street Views on Mood to collect participants’ rating of each street view image with respect to the six mood state indicators, which were motivated, happy, positive-social emotion, focused, relaxed, and

Fig. 2. Research Framework.
depressed (Fig. 3). Regarding demographic data, we only collected age and sex data as per relevant studies, which only collected age, sex, and location data and showed that individual perceptions of urban appearance may not be driven by demographic characteristics (Dubey et al., 2016a; Salesses et al., 2013). Because our all participants were from Guangzhou, we didn’t collect data on their locations. A total of 2,996 sample images were manually selected from the study area to build the dataset for scoring. Images with the best quality (e.g., similar in terms of weather conditions, sharpness, and contrast) were sampled densely and randomly. In this phase, the paired comparisons method was used for scoring the images because this approach is highly practical when studying subjective visual attributes (Dubey et al., 2016a; Zhang et al., 2018). Regardless of sample types, this method evaluates the relative position of a single sample compared with others in a group. Participants can also give more objective, fair, and accurate evaluations through comparisons than through other methods.

As inspired by the biophilia hypothesis, the savannah preference, the attention restoration theory, the stress reduction theory, and other theories and mechanisms regarding nature and human health, the effects of the urban environment on mood can be categorized into emotion, attention, and stress (Glasgow et al., 2019; Jiang et al., 2021). Based on the above, we utilized the six indicators referring to various psychological scales of mood states commonly used in environment and human health research (Roberts et al., 2019). For the emotion dimension, the motivated, happy, positive-social emotion indicators were extracted or summarized from the travel mood scale, profile of mood states, positive and negative affect schedule, mood adjective check list, and Zuckerman inventory of personal reactions scale. “Positive-social emotion” refers to the emotions tied to interpersonal interactions (Koush et al., 2019; Mercer, 2014) such as being sociable, not anxious, or not embarrassed. For the attention dimension, the “focused”
indicator was summarized from the perceived restorative scale and the multi-dimensional mood
questionnaire, representing whether the environment contains interesting features that can
prevent people from getting distracted or help them recover from mental fatigue. For the stress
dimension, the “relaxed” and “depressed” indicators were extracted from the travel mood scale
and the total mood disturbance scale. Therefore, the images were scored based on responses
regarding which place looks more motivated/happier/more likely to arouse positive-social
emotion/more focused/more relaxed/more depressed/? The six questions were asked randomly
alongside paired images and participants were required to choose one of three response options,
manly, that the image above or below was better, or that both images were equal.

The first page of the data collection platform informed participants that all information
would only be used for research and kept anonymous (Fig 3). Targeted participants were aged
between 18 and 65 years old and must have lived in Guangzhou for at least six months. During
the scoring, participants could not see the images’ geographical information. Online
questionnaires were disseminated via WeChat, a communication software popular in China.
Finally, 101,458 paired comparison datapoints were collected from 1,590 online participants in
December 2020. The average age of all participants was 24.52. Most participants in this study
are female (see Table 1), who may be more willing to participate in online surveys than males
(Smith, 2008).
Fig. 3. The user interface of the Assessing the Effects of Street Views on Mood online platform (http://streetview.scurbanlab.com).

Table 1 Paired comparison statistics.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Male (%)</th>
<th>Female (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivated</td>
<td>17,394</td>
<td>30.1</td>
<td>69.9</td>
</tr>
<tr>
<td>Happy</td>
<td>16,661</td>
<td>30.7</td>
<td>69.3</td>
</tr>
<tr>
<td>Positive-social emotion</td>
<td>16,987</td>
<td>29.9</td>
<td>70.1</td>
</tr>
<tr>
<td>Focused</td>
<td>16,836</td>
<td>30.3</td>
<td>69.7</td>
</tr>
<tr>
<td>Relaxed</td>
<td>16,840</td>
<td>30.1</td>
<td>69.9</td>
</tr>
<tr>
<td>Depressed</td>
<td>16,740</td>
<td>30.4</td>
<td>69.6</td>
</tr>
<tr>
<td>Total</td>
<td>101,458</td>
<td>30.3</td>
<td>69.7</td>
</tr>
</tbody>
</table>

3.4. Perceptual mood state score calculation and classification

Based on previous paired comparison results, the scores of the six indicators (corresponding to motivated, happy, positive-social emotion, focused, relaxed, and depressed) for each image were calculated in the dataset $H_m = \{h_{i,m}\}, H_h = \{h_{i,h}\}, H_s = \{h_{i,s}\}, H_f = \{h_{i,f}\},$ $H_r = \{h_{i,r}\},$ and $H_d = \{h_{i,d}\}$ The study applied the strength of schedule (Ordonez and Berg, 2014;
Park and Newman, 2005; Salesses et al., 2013) method to obtain an aggregated evaluation of each image. In this case, the score for each image depends on the likelihood that it will or will not be chosen over the comparison image. Therefore, we defined the likelihood of being selected \(P_{i, q}\) and not being selected \(N_{i, q}\) and the mood state scores \(H_{i, q}\) of image \(i\) concerning indicator \(q \in \{m, h, s, f, r, d\}\) in the following manner:

\[
P_{i, q} = \frac{p_{i, q}}{p_{i, q} + n_{i, q} + e_{i, q}}
\]

\[
N_{i, q} = \frac{n_{i, q}}{p_{i, q} + n_{i, q} + e_{i, q}}
\]

\[
H_{i, q} = \frac{10}{3}(P_{i, q} + \frac{1}{p_{i, q}} \sum_{a=1}^{P_{i, q}} P_{a, q} - \frac{1}{n_{i, q}} \sum_{b=1}^{n_{i, q}} N_{b, q} + 1)
\]

where \(p_{i, q}\), \(n_{i, q}\), and \(e_{i, q}\) refer to the number of times image \(i\) was selected, not selected, or considered equal to its paired image for question \(q\). The constant values \(\frac{10}{3}\) and 1 were used to adjust the score to fall between 0 and 10, following related studies (Nasar, 1990; Ordonez and Berg, 2014; Salesses et al., 2013; Zhang et al., 2018). To distinguish the images with high \(H_{i, q}\) from the images with low \(H_{i, q}\), based on the suggestions of previous studies (Ordonez and Berg, 2014; Zhang et al., 2018), we defined the binary labels \(y_{i, q} \in \{1, -1\}\) as follows:

\[
y_{i, q} = \begin{cases} 
1 & \text{if } H_{i, q} > \mu_q + \delta \times \sigma_q \\
-1 & \text{if } H_{i, q} < \mu_q - \delta \times \sigma_q 
\end{cases}
\]

where \(\mu_q\), \(\delta\), and \(\sigma_q\) refer to the mean value, the ratio variable, and the standard deviation, respectively, of the dataset for indicator \(q\). The ratio variable \(\delta\) determines the sampling threshold. The two thresholds, \(\mu_q + \delta \times \sigma_q\) and \(\mu_q - \delta \times \sigma_q\), created an obvious classification between high...
and low $H_{i,q}$. The samples with middle values were neglected because the way they were categorized may not be stable across individuals (Ordonez and Berg, 2014). The $\delta$ was then adjusted to train the prediction model.

Following a previous study (Salesses et al., 2013), we calculated the inter-rater robustness and internal consistency of the H-scores to validate the process of scoring. The inter-rater robustness was measured by using the average R-square of the Pearson correlation between H-scores calculated using the same number of images but extracted from disjoint subsets of votes of size. From experience, we found that each image needed to appear between 22 and 30 times for each of the six questions to achieve inter-rater robustness greater than 75%. However, we can only collect 11.17 comparisons per image per question on average, which did not meet expectations but the results are still instructive according to a previous study (Dubey et al., 2016b). The internal consistency of H-scores was tested by transitivity which was high in our research. The overall level of transitivity of our data was 81.80\% for “motivated,” 86.92\% for “happy,” 73.53\% for “positive-social emotion,” 73.71\% for “focused,” 87.61\% for “relaxed,” and 87.18\% for “depressed”.

To explore potential sample biases, we compared the correlations of the H-scores obtained for participants younger than the median age of 24.52 to those of participants older than the median age, and we compared the correlations of the H-scores obtained for male participants to those of female participants, with the correlations obtained for the two disjoint random half-samples of participants. We also calculated the correlations obtained for random subsets of participants to those of control groups of the same size. Finally, we found that the correlations obtained for different groups of people are not significantly lower than those obtained for the random control
groups (see Appendix B). Overall, results indicated that no significant bias for different age or gender groups.

3.5. Image segmentation

To predict the mood states in large urban areas and identify the effects of street view elements on mood, we extracted street view elements with PSPNet trained on Cityscapes to compute the percentage of sky, greenery, buildings, and so on in each image. PSPNet is a scene parsing model based on a deep convolutional neural network with an accuracy of 80.2% in identifying 19 categories of objects on Cityscapes (Zhao et al., 2017). Cityscapes is a widely used dataset that can adequately capture the complexity of real-world urban scenes and train and test processing for pixel-level and instance-level semantic labeling (Cordts et al., 2016). Table 2 lists the descriptive statistics of the street view features extracted from the images by semantic segmentation. Ten categories of objects with high frequency and good quality results were selected for the study. As shown, roads, vegetation, buildings, and the sky are the most frequent elements in this study.

Table 2 Descriptive statistics of the street view features extracted by semantic segmentation.

<table>
<thead>
<tr>
<th>Label</th>
<th>Number of images</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road</td>
<td>195,299</td>
<td>0.26</td>
<td>0.08</td>
<td>0</td>
<td>0.53</td>
</tr>
<tr>
<td>Vegetation</td>
<td>195,159</td>
<td>0.24</td>
<td>0.15</td>
<td>0</td>
<td>0.9</td>
</tr>
<tr>
<td>Building</td>
<td>193,581</td>
<td>0.17</td>
<td>0.15</td>
<td>0</td>
<td>0.99</td>
</tr>
<tr>
<td>Sky</td>
<td>193,002</td>
<td>0.15</td>
<td>0.11</td>
<td>0</td>
<td>0.49</td>
</tr>
<tr>
<td>Car</td>
<td>192,336</td>
<td>0.04</td>
<td>0.05</td>
<td>0</td>
<td>0.95</td>
</tr>
<tr>
<td>Sidewalk</td>
<td>193,626</td>
<td>0.03</td>
<td>0.03</td>
<td>0</td>
<td>0.37</td>
</tr>
<tr>
<td>Fence</td>
<td>188,787</td>
<td>0.03</td>
<td>0.04</td>
<td>0</td>
<td>0.69</td>
</tr>
<tr>
<td>Terrain</td>
<td>182,922</td>
<td>0.03</td>
<td>0.04</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>Wall</td>
<td>182,963</td>
<td>0.02</td>
<td>0.03</td>
<td>0</td>
<td>0.56</td>
</tr>
<tr>
<td>Truck</td>
<td>47,454</td>
<td>0.00</td>
<td>0.02</td>
<td>0</td>
<td>0.53</td>
</tr>
</tbody>
</table>
3.6. Modeling the impact of street environment on mood states in a large urban area

According to previous studies (Ordonez and Berg, 2014; Zhang et al., 2018), it is practical to consider the modeling task for mood states in a large urban area as a binary classification task when studying subjective visual attributes. To ensure the quality of the modeling and prediction results of different mood states, we first compared the performance of the five machine learning algorithm models to one another after training them to separately predict the six indicator scores separately. Namely, the machine learning models used were K-nearest neighbor (KNN), random forest (RF), naïve Bayes (NB), logistic regression (LR) and support vector machine (SVM). Converting the complex prediction task into a binary problem minimized the noise of the prediction. Inspired by previous studies (Ordonez and Berg, 2014; Zhang et al., 2018), we selected representative samples from the 2,996 images with H-scores and classified them as high-score samples ($y_{l,q} = 1$) and low-score samples ($y_{l,q} = -1$) for the learning and testing task. The typical method used to evaluate the predictive quality is to split the sample dataset, $S$, into a learning set, $A$, and a testing set, $B$, such that $S = A \cup B$ and $A \cap B = \emptyset$.

The sample dataset was randomly split into 70% for the learning network and 30% for testing the model accuracy and generalizability. Among the five machine learning algorithms, SVM showed the most stable performance (Table 3). Thus, we ultimately chose a supervised learning technique, SVM, to train a classification model, which performed well in judgment tasks. SVM is a binary classifier that explores the widest margin in the classification boundaries to reduce the uncertainty of discrimination (Jahani et al., 2020). The SVM algorithm equation can be represented as follows:

$$f(x) = sgn \left( \sum_{i=1}^{N} a_i y_i K(x_i, x_j) + b \right)$$
where $x_i$ and $x_j$ are the training and test patterns, respectively; $i=1,2,3..., n$; $a_i$ is the optimal solution of the quadratic function; $y_i \in \{1, -1\}$ are the class labels; $K(x_i, x_j)$ represents the kernel function; and $b$ denotes a classification threshold. The most common basis function, the Gaussian radial basis function (RBF), was used in this case, which is defined as follows:

$$K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)$$

We then applied the pretrained model to predict the mood states elicited by images of Guangzhou's central districts (i.e., Tianhe, Haizhu, Yuexiu, and Liwan). To achieve a comprehensive evaluation of mood states, we used the previously collected 199,754 geotagged street view images from Tencent Street View from throughout the study area. As the results of the SVM model were binary, we employed positive confidence, which indicates the predictive probability of one sample being a high-score sample ($f(x) = 1$), to map the effects of the urban street on mood states.

Table 3 Comparison between different machine learning algorithms (average accuracy ± standard deviation).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Motivated</th>
<th>Happy</th>
<th>Positive-social emotion</th>
<th>Focused</th>
<th>Relaxed</th>
<th>Depressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>0.74 ± 0.06</td>
<td>0.76 ± 0.05</td>
<td>0.70 ± 0.05</td>
<td>0.70 ± 0.06</td>
<td>0.78 ± 0.06</td>
<td>0.75 ± 0.05</td>
</tr>
<tr>
<td>RF</td>
<td>0.77 ± 0.03</td>
<td>0.80 ± 0.04</td>
<td>0.72 ± 0.05</td>
<td>0.73 ± 0.05</td>
<td>0.81 ± 0.05</td>
<td>0.77 ± 0.06</td>
</tr>
<tr>
<td>NB</td>
<td>0.70 ± 0.07</td>
<td>0.75 ± 0.06</td>
<td>0.62 ± 0.08</td>
<td>0.70 ± 0.05</td>
<td>0.76 ± 0.05</td>
<td>0.73 ± 0.05</td>
</tr>
<tr>
<td>LR</td>
<td>0.76 ± 0.04</td>
<td>0.79 ± 0.04</td>
<td>0.72 ± 0.06</td>
<td>0.75 ± 0.04</td>
<td>0.82 ± 0.04</td>
<td>0.77 ± 0.06</td>
</tr>
<tr>
<td>SVM</td>
<td>0.76 ± 0.04</td>
<td>0.79 ± 0.05</td>
<td>0.71 ± 0.05</td>
<td>0.75 ± 0.04</td>
<td>0.82 ± 0.04</td>
<td>0.77 ± 0.05</td>
</tr>
</tbody>
</table>

Note: KNN — K-nearest neighbor, RF — random forest, NB — naïve Bayes, LR — logistic regression, SVM — support vector machine.
3.7. Assessing different street view elements effect on mood

To assess the effects of different street view elements on mood, bivariate correlation analysis (Pearson correlation analysis) among the variables and three feature selection techniques (i.e. stepwise regression, ridge regression, and lasso regression) were used separately for each of the six mood state indicators (Fig. 4). Stepwise regression assumes that the relationship between mood and the predictors is linear, and it is a subset selection technique that identifies the optimal subset of predictors. Lasso regression and ridge regression are regularization or shrinkage algorithms that use similar shrinkage penalties to fit the model. Broadly, lasso regression selects essential variables by shrinking some coefficient parameters to zero, while ridge regression keeps all variables in the final models. According to these methods, we calculated the number of times that the street view elements were identified as important predictors, which reflects the robustness of the predictors. Predictors’ estimates and their directions, based on three feature selection techniques for six indicators, were also presented and discussed to assess the effect of street view elements on mood. We used the data from the online platform Assessing the Effects of Street Views on Mood for analysis. For these image data, H-scores were calculated, and the area percentage of the street elements was also computed. Specifically, each of the six mood state indicators were used as a dependent variable, and the percentages of sky, greenery, buildings, and so on were used as the predictor variables.
4. Results

4.1. Prediction experiment results

For the SVM model to achieve high-quality prediction performance, it was trained and cross-validated. The ratio variable $\delta$ was adjusted for the model training performance. Fig. 5 presents the number of samples selected and the accuracy of the model with different $\delta$ values, namely, 0.1, 0.3, 0.5, 0.7, 1.0, 1.2, 1.5, and 1.8. The gap between the high-score and low-score images depended on the $\delta$ value, which also determined the sample size used for testing. As $\delta$ decreased, more samples were selected. A k-fold cross-validation method was used to estimate prediction accuracy (Fushiki, 2011). This technique was used to evaluate the model by dividing the original sample randomly into k subsamples of equal-size. One subsample is retained as a test set, while the remaining k-1 subsamples are used as training sets, and the process then repeats k times. In this stage, a k value of 10 was used for the study. The mean accuracy increased as the ratio variable $\delta$ increased, indicating that the sample size decreased (Fig. 5). We also adjusted penalty coefficients to obtain the highest possible accuracy (see Appendix A). Generally, the
higher the penalty coefficient is, the higher the accuracy is, but the generalizability weakens. The models achieved high accuracy and retained a relatively strong generalizability when the penalty coefficient was equal to 1 that we this applied in final models.

The results yield several conclusions. First, the SVM models with different $\delta$ values demonstrate good performance, with high and reliable accuracy (>60%) for the six mood state indicators. Second, classification is easier when $\delta$ is larger but the sample size is smaller. We ultimately used the model with $\delta$ equal to 1.0, which had an accuracy of over 70% for the prediction of all indicators. Third, as $\delta$ changes, the accuracy of positive-social emotion changed slightly less than that of other indicators. Thus, positive-social emotion seems difficult to evaluate and predict. This might be because individuals’ definition of positive-social emotion is unstable for street view images.
Fig. 5. (a) The bars show the different high-score (1) and low-score (-1) sample sizes for the six indicators used in the training task with different values of δ. More samples were selected as δ decreased. (b) The six curves illustrate the mean accuracy of the SVM model.

4.2. Mapping mood in an urban-scale street environment

Based on the prediction experiment, we applied the SVM model to street view images of the study area. We calculated the average positive confidence values of street images captured in four directions at one location point. The quantified mood states, which are represented by the value of positive confidence values, were color-coded for visualization based on geographical coordinates from QGIS across the study area.
Fig. 6 shows the study area’s distribution maps of mood states for the six different indicators. Overall, some of the mood state indicators are distributed in a similar spatial pattern. The urban fringe areas are more likely to elicit more motivated, happy, focused, and relaxed feelings in the residents than the city center area. Conversely, a more depressive area is concentrated in the city center, covering most of Yuexiu district, the southwest part of Haizhu district, and southern part of the Tianhe district (see Fig. 6f). The city center area of Guangzhou is a built environment in which commercial and residential functions are concentrated, while several large city parks, which provide more green spaces viewable from the surrounding streets, are located in the urban fringe area. This indicates that there may be a correlation between mood and the city’s actual landscape characteristics. This finding is consistent with previous works suggesting that urban environments inhibit restoration potential, and, conversely, that natural environments improve cognitive restoration and increase positive moods through stress recovery (Kaplan, 1995). Interestingly, in terms of the “positive-social emotion” indicator, streets surrounding residential areas generally score higher than arterial roads. Moreover, it is notable that streets along blue spaces make people feel more motivated, happy, focused, and relaxed, as well as less depressed.
Fig. 6. Mapping the effects of urban streets on the six mood state indicators.

4.3. Street view elements’ effect on mood

Fig. 7 displays the correlation matrix of the variables, including the mood state indicators and street view elements. Generally, most of the street view element variables show weak correlations with one another, with correlation coefficients below 0.5. Most of the street view
element variables are negatively correlated with the five positive mood state indicators, while “vegetation” and “terrain” are positively correlated with the five positive mood state indicators and not with the “depressed” indicator. Moreover, of the six mood state indicators, results indicate that the five positive mood state indicators are positively correlated with one another. By contrast, the five positive mood state indicators are negatively correlated with the “depressed” indicator.

To determine the effects of different street view elements on mood, three feature selection techniques, stepwise regression, ridge regression, and lasso regression, were used. Fig. 8 presents the number of times (ranging from 0 to 3) that the street view element variable was identified as a robust predictor. Notably, as shown in Fig. 8, “vegetation” and “terrain” are the two most important variables, with the highest explanatory power across all six mood state indicators. As presented in Fig. 7 and Fig. 9, “vegetation” and “terrain” are both positively correlated with five positive mood state indicators and negatively correlated with the “depressed” indicator. Moreover, according to Fig. 7 and Fig. 9, we can conclude that different street view elements affect the six indicators differently. For instance, “road” is a robust indicator positively correlated with the “motivated” indicator and negatively correlated with the “depressed” indicator, but no correlation was found between “road” and other mood state indicators. “Sky” is also negatively associated with the “positive-social emotion” and the “depressed” indicators but positively associated with the “motivated” indicator. “Building” is a robust and negative predictor for the “focused” and “happy” indicators, and it is positively related to the “depressed” indicator.

To further verify the effects of street view elements on mood using the above techniques, images with low, medium, and high H-scores based on the manual scoring process were
extracted (Fig. 10). For instance, images with higher “motivated,” “happy,” and “relaxed” scores have a higher proportion of vegetation and more diverse terrain, while images with higher “depressed” scores are darker and display a lower proportion of sky. More “focused” images have fewer buildings, and images that scored higher for “positive-social emotion” images have a higher proportion of vegetation, more diverse terrain, more sidewalks, and less sky. These results align with the findings estimated via the regression techniques.

Fig. 7. Correlation matrix of the variables. The color indicates the magnitude of the coefficients between the variables (***p<0.001, **p<0.01, *p<0.05).
Fig. 8. Street view elements that affect mood. A darker color implies that the predictor has a more robust explanatory power.
Fig. 9. Predictors' estimates and their directions based on stepwise regression, ridge regression, and lasso regression for six mood states. (**p<0.001, *p<0.05).
Fig. 10. Images sampled from Guangzhou with different H-scores based on the manual scoring process.

5. Discussion

5.1. Association between urban streets and mood

There is convergent evidence that well-designed urban environments contribute to shaping residents’ moods (Dong and Qin, 2017; Helbich, 2018). Most studies that examine the impact of physical urban environments on mental health do so through participants’ scoring or GIS-derived measures (Nordbø et al., 2018; van den Berg et al., 2015). Recently, a number of studies have demonstrated the benefit of using street view data sources to capture perceptions of an urban environment using similar approaches (Ramírez et al., 2021; Rossetti et al., 2019; Zhang et al., 2018). However, there is a paucity of street view data characterizing residents’ moods. This study proposes a novel methodology for measuring the effects of urban streets on
mood using street view images in large urban area in conjunction with machine learning techniques. Another important contribution of this research is the exploration of the impacts of street view elements on residents’ mood.

Applying this approach to measure the spatial distribution of mood states in Guangzhou, we found that mood states vary geographically at a city scale and that some of the mood state indicators are distributed in a similar spatial pattern. With a high-rise, high-density landscape and less greenery, the city center area induces stress in residents and triggers more depressive feelings. In contrast, the less developed peripheral areas obtained good mood state scores due to sufficient greenery on streets and fewer tower blocks. This aligns with previous literatures indicating a dose-response relationship between greenery and emotion (Chan et al., 2021).

Living streets play a crucial role in social interaction, which coincides with the literature suggesting that streets are not only pathways for travel but also vital public areas for interpersonal interactions among neighbors (Jacobs, 1961), these interactions, in turn, may decrease the negative moods. Results also indicate that streets with blue spaces (e.g., rivers and lakes) have a positive effect on mental feeling effect. These findings correspond to the findings of most environmental psychology researchers, who have argued that blue spaces, meaning outdoor environments that prominently feature water, may improve better mental states and reduced levels of depression (Grellier et al., 2017; Subiza-Pérez et al., 2020; Vert et al., 2020).

The mappings of mood states based on street view elements have implications for improving cities’ physical characteristics, and they can inform public policy and future urban designs of healthy cities.

Consistent with previous studies (Chan et al., 2021), our findings also revealed weak but significant correlations between street view elements and mood. The relationships between mood
and street view elements may be indirectly mediated through various mechanisms (Lopes et al., 2020). Additionally, based on statistical analysis of feature selection, results indicate that the variables “vegetation” and “terrain” have a robust and significantly positive impact on all positive mood state indicators but a negative impact on the “depressed” indicator. Numerous studies have confirmed that vegetation exerts beneficial effects on mood. For instance, Jiang et al. (2014) found that a 1.7% to 24% increase in tree density may result in an increased stress recovery for men. However, an American study showed that increased tree canopies are related to stress reduction, while understory vegetation negatively impacts mental health (Jiang et al., 2020). This result aligns with Norman (Norman K. Booth, 2011) and Gehl (Gehl., 2011), who suggested that the spaces formed by vertical changes in terrain elicit different moods in different groups. In addition, a study from Nanjing, China, indicated that street pavement slope was significantly related to physical activity, which also indicated that terrain may be indirectly linked with people’s moods (Wu et al., 2019).

Interestingly, “road” is positively associated with the “motivated” indicator and negatively correlated with the “depressed” indicator. On the one hand, we found that images with higher H-scores for the “motivated” indicator often show roads, vegetation, and other elements with vibrant colors such as red and yellow. Similar to Bellizzi and Hite (1992), and Blijlevens et al. (2012), we found that warm and highly saturated colors have a physiologically uplifting effect on people. On the other hand, viewing urban roadways or concrete surfaces can increase people’s negative feelings such as depression (Huang et al., 2020; Wang et al., 2016). In addition, “sky” was found to be negatively related to the indicators “positive-social emotion” and “depressed,” but it was positively associated with the “motivated” indicator. This may be because crowded and dense neighborhoods with low sky visibility are likely to be more conducive to the
construction of social networks than other neighborhoods (Wang et al., 2020), while spaces that are too narrow probably have a negative effect on mood. “Building” is also negatively correlated with some positive mood states while positively related to the “depressed” indicator, which aligns with prior studies. Denser and taller buildings may result in an unfavorable microclimate environment and be more oppressive, thus affecting individuals’ mood (Meng et al., 2020; Wang et al., 2020; Zarghami et al., 2019). Additionally, the “happy” indicator is robustly and negatively related to “truck,” “car,” and “wall.” The “positive-social emotion” indicator is robustly and positively correlated with “sidewalk” but negatively related to “truck.” This can be explained by the fact that congested, and chaotic traffic and enclosed environments may make people feel nervous and unhappy. “Sidewalk” is probably related to perceived walkability, which may be beneficial in alleviating negative social emotions such as anxiety, shame, embarrassment, and jealousy (Chen et al., 2020; Wang et al., 2019). However, results indicate that “fence” may have an insignificant relationship with mood. It can be argued that using street view images to identify details related to certain moods may not be reliable (Kang et al., 2020). In this study, the proportion of the “fence” element was relatively small in most images (Table 2).

5.2. Limitations and future studies

The following limitations of this study should be noted. First, our research was cross-sectional, which prevented us from drawing conclusions regarding causality among the variables assessed. Thus, longitudinal or experimental studies should be considered in future works.

Second, the quality and variation of images in the dataset and the performance of the computer visualization models may affect the accuracy of results (Kang et al., 2020; Verma et al., 2020). Although PSPNet achieved a state-of-the-art performance, segmentation of visual
scenes executed by PSPNet model trained on Cityscapes can only automatically identify 19 categories of semantic objects. Detailed information is lost in the segmentation process because features of the images that are not included in the 19 Cityscapes’ classifications may be falsely categorized as one semantic object. Methodological developments in fine-grained image extraction are needed to further advance related research. Mapping companies and researchers in computer visualization and machine learning could improve the dataset and the algorithms. This would enhance the results of future studies.

Third, although previous studies have demonstrated the efficacy of using street view images to observe the relationship between perceptions and the urban environment, several limitations of using street view images also warrant discussion. Street view images cannot capture changes in factors such as time and weather because they are not updated frequently (Kang et al., 2020). In addition, human emotion is impacted by factors beyond visualization (Zhang et al., 2018). Therefore, we cannot be sure that a singular viewpoint limited to certain times is representative of a place. More techniques, such as videos, virtual environments, and 3D reconstruction of scenes, should be used in future studies (Ramírez et al., 2021). Another possible bias may be introduced by the fact that street view images are taken from the perspective of a vehicle because these images might not accurately represent the streetscape view of pedestrians and bicyclists on the sidewalks or cycling paths (He et al., 2020; Verma et al., 2019). Future studies should try to collect the streetscape photos from pedestrian or bicyclists centered perspectives, which might more accurately represent the perceived environment of populations with different transport modes or preferences. Moreover, Tencent street view images can only show the environment along major roads and cannot be applied to parks, in
neighborhoods, campuses, and so on. Future studies could investigate these places where people’s activities often take place.

Fourth, we only considered several semantic segmentation elements of street view images as predictor variables in the regression models and did not consider effects brought about by other environmental elements and spatial autocorrelations. The natural and artificial elements extracted from the street view images may not constitute a complete street environment. Many other street features could affect residents’ mental feelings, such as street sanitation, road age, infrastructure configuration, building density, building height, street or ground green perception, and so on (Nordbø et al., 2018; L. Zhang et al., 2019). Further studies should take the above factors into account and combine multiple sources of urban data, such as points of interest, street view images, remote-sensing data, location-based service positioning data, to explore the impact of other street features on residents’ moods (Liu et al., 2020; Tang et al., 2020). They should also apply more analyses techniques, such as linear mixed model analyses, in consideration of the spatial autocorrelations (Zhang et al., 2018).

Fifth, the study adopted a convenience sampling method through WeChat, therefore, the respondents skewed younger. WeChat users tended to share the survey with their friends, classmates, colleagues, and families, which may have led to sample bias. Images were scored by participants with better access to the Internet than other groups (e.g., the elderly or people with lower incomes) (Apuke and Iyendo, 2018), so their perceptions may not represent those of all residents. Moreover, we did not investigate the impact of street environment on mood for different populations. However, perceptions or emotions associated with urban streets may vary among various groups. For example, employees who commute, children who attend school, and retired people who do not leave home may perceive the street environment differently. In
addition, random or unreliable responses provided by the participants may occur in online
surveys, which will increase errors in the training data. In the future, more survey methods can
be applied, such as field investigations, to randomly recruit more participants from different
contexts and collect more detailed demographics information (e.g., education and income).
Socioeconomic factors must be considered as input to help test individual emotions associated
with the street environment in the predicting model.

Sixth, we evaluated mood states based on the images, which respondents voted on an
online platform through pairwise comparisons, to obtain a large sample size, and we used this
method to assess mood states in a large urban area. However, our results may be biased because
there may be a difference between our online assessment and respondents’ actual exposure to
real environment. The online street images lack non-visual sensory information and may not
elicit the same moods of real-world exposure would. Mobile sampling designs utilizing wearable
devices in the field should be combined with online measures in future research to explore the
difference between emotion elicited by street view images and by the place in-situ. Meanwhile,
although pairwise comparisons are beneficial for people to make a more impartial assessment, it
has the limitation of requiring a large amount of data to achieve a stable ranking. More
techniques, such as human-machine adversarial methods (Dai et al., 2021), could be applied in
future studies to ensure accuracy and save time and labor in the scoring process.

6. Conclusion
Exposure to the urban environment can affect various aspects of the residents’ health and
well-being. However, there is a paucity of crowdsourcing platforms based on street view data,
which can be used to characterize residents’ moods. This study proposes a novel methodology to
measure the effects of urban streets on mood states, using street view images at a massive scale
and machine learning techniques. Specifically, using the case of Guangzhou, this study used a trained machine learning model to predict and map qualitative mood states based on local street view images. Furthermore, correlation analysis and three feature selection techniques were used to assess the impact of street view elements on mood.

The SVM model shows good performance with high accuracy for six mood state indicators. The maps of the predictive effects of urban streets on mood four districts of Guangzhou, China, show that streets in urban fringe areas are more likely to produce more positive feelings in residents than those in city center areas. Conversely, the city center, a high-density built environment, is more likely to produce depressive feelings in residents. The results also indicate that the variables “vegetation” and “terrain” may have significant and positive effects on all mood state indicators, namely, motivated, happy, positive-social emotion, focused, relaxed, and depressed. Our findings could prompt urban designers to consider which areas of the city are crucial for optimization, and they have practical implications for planning healthy cities planning. This study is a preliminary attempt to provide an improved research methodology for modelling the effects of urban streets on mood. The results demonstrate that street view data can reveal mood state patterns at a city scale and can serve as a fine-granularity measure of street landscape factors.

Reference


https://doi.org/https://doi.org/10.1002/psp.1892


https://doi.org/https://doi.org/10.1016/S0272-4944(02)00109-3


https://doi.org/https://doi.org/10.1016/j.envres.2017.11.006


Li, X., Zhang, C., Li, W., 2015. Does the visibility of greenenery increase perceived safety in urban areas? Evidence from the place pulse 1.0 dataset. ISPRS Int. J. Geo-Information. https://doi.org/10.3390/ijgi4031166


