An unsupervised machine learning approach to evaluate sports facilities condition in primary school

Jun Xia¹, Jihong Wang¹, Hua Chen², Jie Zhuang¹, Zhenbo Cao¹, Peijie Chen¹*

¹ School of Kinesiology, Shanghai University of Sport, Shanghai, China, ² Institute of Urban Development and Management, Tongji University, Shanghai, China

* chenpeijie@sus.edu.cn

Abstract

Sports facilities have been acknowledged as one of the crucial environmental factors for children’s physical education, physical fitness, and participation in physical activity. Finding a solution for the effective and objective evaluation of the condition of sports facilities in schools (SSFs) with the responding quantitative magnitude is an uncertain task. This paper describes the utilization of an unsupervised machine learning method to objectively evaluate the condition of sports facilities in primary school (PSSFC). The statistical data of 845 samples with nine PSSFC indicators (indoor and outdoor included) were collected from the Sixth National Sports Facility Census in mainland China (NSFC), an official nationwide quinquennial census. The Fuzzy C-means (FCM) algorithm was applied to cluster the samples in accordance with the similarity of PSSFC. The clustered data were visualized by using t-stochastic neighbor embedding (t-SNE). The statistics results showed that the application of t-SNE and FCM led to the acceptable performance of clustering SSFs data into three types with differences in PSSFC. The effects of school category, location factors, and the interaction on PSSFC were analyzed by two-way analysis of covariance, which indicated that regional PSSFC has geographical and typological characteristics: schools in the suburbs are superior to those in the inner city, schools with more grades of students are configured with better variety and larger size of sports facilities. In conclusion, we have developed a combinatorial machine learning clustering approach that is suitable for objective evaluation on PSSFC and indicates its characteristics.

Introduction

The prosperity of a country, to some extent, relies on the comprehensive literacy of the young generation, among which physical fitness (PF) is regarded as one of the fundamental and vital elements [1]. Sufficiency of physical activity (PA) at a moderate to vigorous level, recommended by WHO [2], is a well-researched contribution to facilitate health-related PF and reduction of childhood obesity, non-communicable chronic diseases, heart disease, and diabetes [3–7]. In general, schoolchildren’s PF and PA are diversified not only by individual
characteristics [8] but also by environmental factors [9–11]. Among the latter, in contrast to the family and community, school is widely perceived as the subject of admission and responsibility for the promotion and intervention of schoolchildren’s PA [12]. Previous investigators have identified the positive correlation between the diversity of SSFs and children’s PA [13,14]. Moreover, as the solid foundation of high-quality physical education, the excellence of SSFs motivates schoolchildren to be physically active during recess and lunchtime [15]. The discrepancies in SSFs determine the variance of schoolchildren’s accessibility and availability to school-based PA participation. Therefore, the condition of SSFs can be deemed as the indirect and non-ignorable factor of PA promotion in childhood.

In recent years, several researchers have sought to classify SSFs for the quantification of correlation between differentiated SSFs and students’ PA. Fórnias et al. [16] introduced the “number of PA facilities” variable into the analysis of the relationship between adolescent PA and extracurricular sports activities. This variable was composed of the quantity of six types of SSFs. Lo et al. [17] conducted a cross-sectional study on the Taiwanese adolescent school sports condition and relevant activities in recess time, with the consideration of the investigated schools with or without a school sports field and gymnasium. Haug et al. [18] generated an outdoor facility index (summarized and standardized score of outdoor SSFs numbers) based on a school-level questionnaire for Norwegian primary and secondary schools. The results revealed its correlation with students’ physical activity. Bevans et al. [19] proposed a regression model to evaluate PE effectiveness by interviewing teachers at three school districts in the United States. In this model, school facility resources were rated into three levels, according to teachers’ satisfaction with six types of sports facilities. In essence, existing studies have involved the assessment or classification based on the actual quantity and subjective satisfaction of SSFs. However, few writers have been able to draw on any systematic research into the size factor of SSFs. Meanwhile, extensive researches have been carried out in both developed and developing districts, but the issue in mainland China has rarely been documented. As far as our knowledge, there have been few studies carried out to exert an objective SSFs evaluation for primary schools with high effectiveness and efficiency.

To address the drawback of traditional subjective approaches (users’ comments, questionnaire, interview, grading scales, etc.), the major purpose of this study is aiming to provide a novel combined machine learning method on the feature extraction and evaluation of PSSFC based on the objective SSFs data. Firstly, FCM was utilized to cluster investigated samples with the sample into the same categories. The characteristics of categories in PSSFC are exhibited accordingly, which gives a good understanding of the PSSFC differences within a specific region. Secondly, t-SNE was used to perform the visualization of clustered results in the lower dimension to give an explicit illustration of the distribution. Thirdly, a web-crawling method is used to acquire the individual’s location as the supplementary attribution for further analysis. Based on the above results, the two-way ANOVA analysis was conducted to indicate the impact and interaction of the individual’s location and school category on its PSSFC in the same clustered PSSFC type. The results of the experiment showed that the machine learning approach by FCM with t-SNE in this paper performs with effective and interpretable performance on the aspect of feature extraction and evaluation and on PSSFC.

Methodology

Data collection

This study was based on the Sixth National Sports Facility Census (NSFC, 2014) conducted by the General Administration of Sport of China to survey the quantity, distribution, and purpose of the existing sports facilities in mainland China. The School of Kinesiology at Shanghai
University of Sport has been authorized to access and collect these data via the specific data interface.

**Subjects.** All investigated samples were selected that met all of the following criteria from the NSFC database: (1) school type: public schools rather than private schools, considering the equity of government investment in education; (2) geographic scope: all primary schools in the studied region. The investigated region was Shanghai, which is a well-developed international city and consisted of 7 urban districts, 8 suburban districts, and 1 urban-suburban fringe district. The reason is that the primary educational resources and infrastructure in the city, conspicuously exhibit regional characteristics in suburban and urban districts, owing to the imbalance of population and economic development. Thus, the distribution and condition of SSFs in the city could be deemed as the epitome of the world. In detail, stipulated by the Law of the People’s Republic of China on Education, the subjects were allocated into three school categories (see Table 1): primary school (5-year curriculum, PS, n = 666), nine-year schools (Grade 1 to 9 combined, 9CS, n = 150) and twelve-year schools (Grade 1 to 12 combined, 12CS, n = 29). The inequalities in the distribution of samples emerged because all schools in the investigated region were studied in this paper. As for the geographical distribution pattern, the urban primary schools and the suburban were approximately equal in quantity with ratios of 1:1.07.

**School facility condition measurements.** The statistical indicators of NSFC covered all aspects of sports facilities and were divided into 3 categories: (1) general information (ownership type, ownership details, address, and athletic team service information), (2) sports facility information (type, scale, number of staff, location type, year built, and open service), and (3) operational information (operation mode, hosted sports events, fitness training programs, operation income, and costs). The operational information referred to the utilization and operation of a particular school facility and relevant business. However, sports facilities in all samples in this paper were merely used to meet the demand of primary education for public benefit. Therefore, the operational information was insignificant to this paper and omitted, whereas the rest relating to this study were selected and categorized. The details of each descriptive measurement are described as follows:

(1) Variety. The variety was assessed by the aggregated numbers of sports facility types in each school. In the statistic caliber of NSFC, the types of SSFs were investigated listed in Table 2. The types of indoor, outdoor, and total facilities were computed separately, for the sake of describing the condition of SSFs precisely. Hence, we selected the types of indoor (IT), outdoor (OT), and total facilities (TT) as the indicators of PSSFC variety.

(2) Size. The size of the SSFs was determined by the area of an individual facility. The area was defined as the effective accessible area used for training, competition, fitness activities, and safety protection. To demonstrate the variance of SSFs, the area of indoor (IA), outdoor (IA), and total (TT) is shown in Table 2.

| Primary School Category | Suburban | Urban | Total |
|-------------------------|----------|-------|-------|
| 12CS                    | 13       | 16    | 29    |
| 9CS                     | 89       | 61    | 150   |
| PS                      | 306      | 360   | 666   |
| Total                   | 408      | 437   | 845   |

12CS = 12-year school; 9CS = 9-year school; PS = 5-year school, primary school only.

https://doi.org/10.1371/journal.pone.0267009.t001
(OA), total facilities (TA), and those average values were adopted as the indicators of PSSFC size.

(3) Location. The location (SL) of an individual school was used to differentiate whether an individual belongs to the urban or suburban category. However, the geographical data of SSFs in NSFC were described in form of addresses, which were descriptive and incapable of directly expressing the geographical attribution of each SSFs. Furthermore, as an exceptional case, the Pudong District (the second largest and unique urban-suburban fringe district in Shanghai) is allocated with the most fundamental education resources (4 times more than the largest Chongming District [20]). As an alternative, the individual’s SL in this paper was extracted via programming from the Baidu Map and calculated by measuring the distance of sampled school to the Building of Shanghai Municipal Peoples’ Government (City Hall) and computed by using the Haversine formula [21]. Thus, we divided all samples into urban (with a distance less than equivalent to 15 km, $N = 385$) and suburban (a distance greater than 15km, $N = 460$) in light of the General Development Plan of Shanghai City.

Data processing and analysis
The general framework of analysis in this paper is depicted in Fig 1. All SSFs data were processed and modeled by using Python (Python 3.7.4 version for Windows, the Python Software Foundation, Delaware, USA). The location of each sample was extracted and consolidated with the PSSFC to structure a 9-dimensional SSFs dataset. The original SSFs dataset contained four numeric segments that featured the PSSFC overall (TT, TA, and Average TA), indoor (IT, IA, and Average IA), outdoor (ON, OA, and average OA), and SL. In common terms, the overall segment was the summation of indoor and outdoor, which performed a linear relation. Inversely, the relation between indoor and outdoor segments exhibited nonlinearity and independence. The min-max normalization was operated to scale the data between 0 and 1.

Clustering is usually applied to distinguish the objects with similar attributes by allocating them into different groups (clusters) [22] when the raw data are unlabeled. Based on the reality that no prior model or criteria for the classification of SSFC has been proposed, the solution shifts to rely solely on the features and similarity of SSFC as the criteria for attribute classification, which falls under the realm of unsupervised clustering. Most clustering algorithms were hard clustering (HC) [23,24]. Actually, in some cases, each data sample could present the belongingness to two or more clusters. Thus, the typical HCM algorithms, such as K-means and Hierarchical behaved with low applicability. Therefore, we introduced the Fuzzy-C Mean (FCM) Algorithm [25] to cluster primary schools by similarity in PSSFC and sort to certain

| Category | Types of School Sports facilities | Types |
|----------|----------------------------------|-------|
| Outdoor  | stadium, track field, soccer pitch (futsal and 7-a-side pitch included), basketball court (3-a-side pitch included), volleyball court, badminton court, tennis court, swimming pool, American football/rugby pitch, hockey pitch, table tennis table, handball field, cricket pitch, gate ball court, baseball field, archery field, boce courts, skating field, skateboarding/roller skating field, and fitness equipment | 20    |
| Indoor   | track field, fitness room, basketball court, volleyball court, handball field, gymnastics room, badminton court, table tennis room, martial arts room, fight event training room, fitness room, yoga room, weightlifting room, fencing room, chess, and card room, bowling room, futsal pitch, tennis court, hockey pitch, archery field, equestrian field, ice hockey field, skating field, curling kettle field, skateboarding/roller skating field, squash room, and gateball room | 27    |

https://doi.org/10.1371/journal.pone.0267009.t002
levels according to the characteristics of each cluster. FCM has been regarded as one of the most popular and feasible clustering algorithms [22] to tackle the problems regarding MRI image segmentation [26–28], sentiment classification [29], power quality recognition [30], fault detection [31], disease detection [32,33]. It should be noted that despite the undoubted success of deep learning clusters in computer vision applications, for instance, DeepCluster [34] and DFuzzy [35], we adopted the traditional clustering method for the reason of simplicity and small scale of numeric PSSFC data.

FCM, generated from the fuzzy theory, optimizes the objective function to obtain the degree of membership (DM) for each sample point to all class centers in the interval 0 to 1 (soft clustering) rather than binary values 0 or 1 (HC). As a result, each sample can be distributed to multiple clusters by computing its DM to each cluster. In this issue, given the embedded SSFs dataset \( X = \{ x_1, x_2, \ldots, x_n \} \) (in the two-dimensional space after the process of t-SNE) and \( j \) clustering centers \( C = \{ c_1, c_2, \ldots, c_j \} \), a \( 2 \times n \) matrix \( U \) can be generated to describe the clustered result:

\[
U = \begin{pmatrix}
  u_{11} & \cdots & u_{1j} \\
  u_{21} & \cdots & u_{2j}
\end{pmatrix},
\]

where \( u_{ij} \) ranging from zero to one is the DM for an individual \( x_i \) belonging to category \( j \), which can be calculated by:

\[
u_{ij} = \frac{1}{\sum_{k=1}^{n} \left( \frac{\| x_j - c_k \|}{\| x_j - c_i \|} \right)^{m}} \frac{\sum_{i=1}^{N} u_{ij}^m x_j}{\sum_{i=1}^{N} u_{ij}^m}, \quad c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_j}{\sum_{i=1}^{N} u_{ij}^m},
\]

where \( m \) is a real number greater than 1. After the \( t \) times iteration, FCM attempts to minimize
the cost function as below:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^m \|x_i - c_j\|^2, 1 \leq m < \infty.$$  \hspace{1cm} (3)

FCM terminates until the following condition is compromised:

$$\max_{j} \{|u_{ij}^{(t+1)} - u_{ij}^{(t)}|\} < \varepsilon,$$  \hspace{1cm} (4)

which implies that the result has achieved a comparative (local or global) optimum, and the DM will not be altered significantly if the iteration continues.

Generally, the numbers of cluster center $c$ in FCM are predetermined parameters [36]. However, a lack of standards or guidance for evaluating PSSFC has been an issue for the current studies. Therefore, the numbers of the cluster center were adjusted from one to ten to explore the best classification of SSFs. As for the configuration of FCM, the fuzzy partition coefficient (FPC) [37] was selected as the validity index of clustering shown in Eq (5), and the other parameters [29] were selected as $m = 1.5$, $t = 1000$, and $\varepsilon = 0.001$.

$$FPC = \frac{\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^2 - 1/c}{1 - 1/c}$$  \hspace{1cm} (5)

To visualize 9-dimensional SSFs data, $t$-stochastic neighbor embedding ($t$-SNE) [38,39], was applied to visualize PSSFC in the lower dimension. The $t$-SNE has been proved as a visualization approach for non-linear high-dimensional data with high efficiency and widely used in bioinformatics [40,41], disease assessment [42], signal processing [43], fault identification [44], and other domains. It should be also clarified that notwithstanding the complexity scaling ($O(n^2)$) of $t$-SNE is higher than PCA and LDA, $t$-SNE has an excellent performance in the visual description of SSFs features in primary schools, on account of the finite numbers of measurements and samples. The configuration of the parameters, the perplexity $\text{perp}$ and times of iteration $t$, is an inevitable issue to be tackled when applying the $t$-SNE algorithm.

The original data were labeled with the different types (Type-1, Type-2, and Type-3) after the processing of FCM and $t$-SNE. The above three types of PSSFC were exhibited with descriptive statistics (means and standard errors, mean ± SE). Furthermore, a two-way ANOVA was performed to figure out the differences in the variety and size in each PSSFC type adjusted for the school category and location of primary schools, with the significance level set at 0.05. All analyses were processed using the Statistical Package for the Social Sciences (SPSS) 22.0 on Windows (IBM, Chicago, IL, USA).

**Results**

**Clustering and visualization of PSSFC**

**Result of unsupervised clustering.** Fig 2 presents the results of FCM clustering on the embedded PSSFC via $t$-SNE. The samples clustered into the same level were marked with the same color. To select the optimal $c$, we compared the FCM clustering effect when $c$ ranged from 2 to 10. It could be recognized from Fig 4 that when $c = 3$, the clustering result conformed to human beings’ intuition. To further prove that $c = 3$ is the optimal clustering center, Fig 5 exhibits the DM for an individual sample to each clustered level when $c = 3$ and FPC reached the summit (FPC = 0.9479).
In Fig 3, we adopted the different colors to represent the DM for a particular sample to different levels. As mentioned above, the DM indicates that the strength of belongingness for an individual sample to a clustering level in fuzzy clustering. The Panel (1) in Fig 3 shows the DM for each PSSFC with the initial sequence, where the $x$-axis is the order number of samples, the $y$-axis is the order number of the cluster center, and the $z$-axis is the DM $u_{ij}$ of sample $i$ belonged to cluster center $j$. For a better demonstration of the feasibility of FCM, the order of the original data is sorted and numbered according to the individual’s corresponding DM to a
Fig 4. Embedding of clustered PSSFC into two dimensions via t-SNE.

https://doi.org/10.1371/journal.pone.0267009.g004

Fig 5. Geographical distribution of PSSFC with three categories in Shanghai. (A) The geographical distribution panorama of PSSFC in Shanghai. (B) The distribution of Type-1 PSSFC in Shanghai. (C) The distribution of Type-2 PSSFC in Shanghai. (D) The distribution of Type-3 PSSFC in Shanghai.

https://doi.org/10.1371/journal.pone.0267009.g005
certain level from the largest to the smallest, while the y-axis and z-axis are unchanged, the x-axis is converted to the sorted sequence. The results show that all samples were explicitly allocated into three independent categories by FCM.

**Visualization of clustered PSSFC in two-dimensional space.** Fig 4 shows the visual results of embedding PSSFC into the two-dimensional space by using t-SNE (perp = 15, t = 5000, and η = 200), the space (y1, y2), for all of the 845 sampled primary schools. Data with dissimilarity were isolated in the embedded space, while the samples of similar attributes were almost assembled into the adjacent area. This hinted that the utilization of t-SNE possesses the effectiveness of visualization on this issue and the capability to observe the global overview of clustered PSSFC visually.

**Descriptive characteristics of PSSFC at different levels**

Table 3 exhibits the description and distribution of PSSFC with 3 types in detail generated by the defuzzification of the original dataset. In general, over half of the sampled schools were categorized to Type-3 (n = 463, 54.79%) while Type-2 of PSSFC was 17.52% (n = 148), and Type-1 schools occupied over a quarter. From the perspective of descriptive statistics, it could be concluded from Table 3 that the three categories showed a gradation in the overall PSSFC, the discernable characteristics of Type-3 schools were the deployment with at least 1 indoor facility in the total 3 sports facilities the diversity, and maximum area of sports facilities on average. In contrast to Type-1, the Type-2 and Type-3 schools showed the scarcity in the indoor facility, which indicated the fewer alternatives in the above schools when it suffered from inclement weather or air contamination. In addition, the individuals in Type-2 possess almost twice as large and quantitative as those in Type-1 in total facilities.

**Geographical and typological characteristics of PSSFC**

**Impact of location and school category on PSSFC.** In this paper, the school location (L) and the school category (SC) were selected as the test factors. The effects of two factors and the

|                          | Type-1       | Type-2       | Type-3       |
|--------------------------|--------------|--------------|--------------|
| **Schools, n (%)**       | 234 (27.69%) | 148 (17.52%) | 463 (54.79%) |
| **Total Facilities**     |              |              |              |
| TT (n)                   | 1.00±0.00    | 2.00±0.0     | 3.57±1.38    +
| TA (m²)                  | 1998.91±1539.69" | 3815.92±2185.39"++x | 6289.52±5030.98"++x |
| Average TA (m²)          | 1907.96±1092.69"++x | 1736.41±1143.34"++x |
| **Indoor Facilities**    |              |              |              |
| IT (n)                   | 0            | 0            | 1.09±0.76+   |
| IA (m²)                  | 0            | 0            | 549.9±585.67"++x |
| Average IA (m²)          | 0            | 0            | 432.4±374.57"++x |
| **Outdoor Facilities**   |              |              |              |
| OT (n)                   | 1.00±0.00    | 2.00±0.0     | 2.48±1.24+   |
| OA (m²)                  | 1998.91±1539.69" | 3815.92±2185.39"++x | 5739.62±4776.73"++x |
| Average AO (m²)          | 1907.96±1092.69"++x | 2397.4±1666.11"++x |

Data are expressed as mean ± SE.
* significant difference between urban and suburban primary schools at the same PSSFC level (p<0.05).
* significant difference among different school categories (PS, 9CS, 12CS) at the same PSSFC level (p<0.05).
* significant interaction of the types and location of schools between the PSSFC variables at the same PSSFC level (p<0.05).

https://doi.org/10.1371/journal.pone.0267009.t003
interactions on PSSFC in intraclass schools were respectively analyzed, and the results of which are marked in Table 3. Table 4 indicates that the school location, the school category, and the interaction of the two had significant interclass effects on the performance of PSSFC when using the two-way ANOVA with a 0.95 significance level.

In Tables 3 and 4, the findings revealed that the individual’s location was the major influence on its PSSFC. Specifically for each clustered type, the suburban Type-1 schools, which were configured with 1 outdoor facility and possessed larger size in PSSFC than the urban of the same type ($F = 5.162, p < 0.05$). In the Type-2 schools with outdoor facilities, the numbers of PSSFC (TT and OT) had no significant relationship with the individual’s location ($p > 0.05$).

Table 4. Relationships between school location and school category and PSSFC.

| Type   | Variables | Source | df  | SS    | M     | F     | p-value |
|--------|-----------|--------|-----|-------|-------|-------|---------|
| Type-1 | TT (n)    | L      | 1   | 0.007 | 0.007 | 1.539 | 0.216   |
|        |           | SC     | 2   | 0.000 | 0.000 | 0.028 | 0.972   |
|        |           | L×SC   | 2   | 0.001 | 0.000 | 0.060 | 0.942   |
|        | TA (m³)   | L      | 1   | 2.22×10⁷ | 2.22×10⁷ | 5.162 | 0.02401 |
|        |           | SC     | 2   | 6.80×10⁶ | 3.40×10⁶ | 0.791 | 0.455   |
|        |           | L×SC   | 2   | 7.54×10⁶ | 3.77×10⁶ | 0.878 | 0.417   |
|        | Average TA (m³) | L | 1 | 1.40×10³ | 1.40×10³ | 5.176 | 0.02382 |
|        |           | SC     | 2   | 6.88×10⁶ | 3.44×10⁶ | 1.268 | 0.283   |
|        |           | L×SC   | 2   | 8.56×10⁶ | 4.28×10⁶ | 1.577 | 0.209   |
|        | OT (n)    | L      | 1   | 0.007 | 0.007 | 1.539 | 0.216   |
|        |           | SC     | 2   | 0.000 | 0.000 | 0.028 | 0.972   |
|        |           | L×SC   | 2   | 0.001 | 0.000 | 0.060 | 0.942   |
|        | OA (m³)   | L      | 1   | 2.22×10⁷ | 2.22×10⁷ | 5.162 | 0.02401 |
|        |           | SC     | 2   | 6.80×10⁶ | 3.40×10⁶ | 0.791 | 0.455   |
|        |           | L×SC   | 2   | 7.54×10⁶ | 3.77×10⁶ | 0.878 | 0.417   |
|        | Average OA (m³) | L | 1 | 1.40×10³ | 1.40×10³ | 5.176 | 0.02382 |
|        |           | SC     | 2   | 6.88×10⁶ | 3.44×10⁶ | 1.268 | 0.283   |
|        |           | L×SC   | 2   | 8.56×10⁶ | 4.28×10⁶ | 1.577 | 0.209   |
| Type-2 | TT (n)    | L      | 1   | 0.005 | 0.005 | 0.718 | 0.398   |
|        |           | SC     | 2   | 0.039 | 0.019 | 2.975 | 0.054   |
|        |           | L×SC   | 2   | 0.021 | 0.010 | 1.588 | 0.208   |
|        | TA (m³)   | L      | 1   | 1.48×10⁸ | 1.48×10⁸ | 36.143 | 1.44×10⁻⁴ |
|        |           | SC     | 2   | 1.18×10⁸ | 5.90×10⁹ | 14.409 | 1.97×10⁻⁶ |
|        |           | L×SC   | 2   | 6.50×10⁷ | 3.25×10⁹ | 7.932 | 5.40×10⁻⁴ |
|        | Average TA (m³) | L | 1 | 3.46×10⁷ | 3.46×10⁷ | 39.285 | 4.01×10⁻⁹ |
|        |           | SC     | 2   | 2.37×10⁷ | 1.19×10⁹ | 13.444 | 4.44×10⁻⁶ |
|        |           | L×SC   | 2   | 1.34×10⁷ | 6.69×10⁹ | 7.581 | 7.40×10⁻⁴ |
|        | OT (n)    | L      | 1   | 0.005 | 0.005 | 0.718 | 0.398   |
|        |           | SC     | 2   | 0.039 | 0.019 | 2.975 | 0.054   |
|        |           | L×SC   | 2   | 0.021 | 0.010 | 1.588 | 0.208   |
|        | OA (m³)   | L      | 1   | 1.48×10⁸ | 1.48×10⁸ | 36.143 | 1.44×10⁻⁴ |
|        |           | SC     | 2   | 1.18×10⁸ | 5.90×10⁹ | 14.409 | 1.97×10⁻⁶ |
|        |           | L×SC   | 2   | 6.50×10⁷ | 3.25×10⁹ | 7.932 | 5.40×10⁻⁴ |
|        | Average OA (m³) | L | 1 | 3.46×10⁷ | 3.46×10⁷ | 39.285 | 4.01×10⁻⁹ |
|        |           | SC     | 2   | 2.37×10⁷ | 1.19×10⁹ | 13.444 | 4.44×10⁻⁶ |
|        |           | L×SC   | 2   | 1.34×10⁷ | 6.69×10⁹ | 7.581 | 7.40×10⁻⁴ |

(Continued)
Table 4. (Continued)

| Type      | Variables | Source | df | SS    | M     | F     | p-value       |
|-----------|-----------|--------|----|-------|-------|-------|---------------|
| Type-C    | TT (n)    | L 1    | 63.336 | 63.336 | 38.275 | 1.37 x 10^-09 |
|           | SC 2      | 52.001 | 26.000 | 15.713 | 2.52 x 10^-07 |
|           | L x SC 2  | 8.565  | 4.283  | 2.588  | 0.076  |
| TA (m²)   | L 1       | 1.69 x 10^9 | 1.69 x 10^9 | 100.465 | 1.69 x 10^-21 |
|           | SC 2      | 1.59 x 10^9 | 7.96 x 10^8  | 47.377  | 2.04 x 10^-19 |
|           | L x SC 2  | 3.39 x 10^8 | 1.70 x 10^8  | 10.085  | 5.18 x 10^-5  |
| Average TA (m²) | L 1 | 5.50 x 10^7 | 5.50 x 10^7 | 64.860 | 7.11 x 10^-45 |
|           | SC 2      | 4.32 x 10^7 | 2.16 x 10^7  | 25.444  | 3.35 x 10^-31 |
|           | L x SC 2  | 1.15 x 10^6 | 5.76 x 10^5  | 0.679   | 0.508  |
| IT (n)    | L 1       | 0.829  | 0.829  | 1.521  | 0.218  |
|           | SC 2      | 12.658 | 6.329  | 11.614 | 1.20 x 10^-05 |
|           | L x SC 2  | 0.746  | 0.373  | 0.685  | 0.505  |
| IA (m²)   | L 1       | 1.20 x 10^6 | 1.20 x 10^6 | 4.203  | 4.09 x 10^-2  |
|           | SC 2      | 2.33 x 10^7 | 1.17 x 10^7  | 40.806  | 5.01 x 10^-47 |
|           | L x SC 2  | 3.39 x 10^6 | 1.70 x 10^6  | 5.936   | 2.85 x 10^-3  |
| Average IA (m²) | L 1 | 5.74 x 10^5 | 5.74 x 10^5 | 4.410  | 3.63 x 10^-2  |
|           | SC 2      | 4.37 x 10^6 | 2.18 x 10^6  | 16.765  | 9.43 x 10^-8  |
|           | L x SC 2  | 2.45 x 10^5 | 1.23 x 10^5  | 0.941   | 0.391  |
| OT (n)    | L 1       | 78.654 | 78.654 | 58.283 | 1.35 x 10^-13 |
|           | SC 2      | 16.130 | 8.065  | 5.976  | 2.74 x 10^-3  |
|           | L x SC 2  | 4.277  | 2.139  | 1.585  | 0.206  |
| OA (m²)   | L 1       | 1.60 x 10^9 | 1.60 x 10^9 | 105.085 | 2.52 x 10^-22 |
|           | SC 2      | 1.27 x 10^9 | 6.33 x 10^8  | 41.599  | 2.56 x 10^-47 |
|           | L x SC 2  | 2.91 x 10^8 | 1.45 x 10^8  | 9.552   | 8.64 x 10^-5  |
| Average OA (m²) | L 1 | 6.43 x 10^7 | 6.43 x 10^7 | 28.261  | 1.66 x 10^-7  |
|           | SC 2      | 8.09 x 10^7 | 4.04 x 10^7  | 17.763  | 3.73 x 10^-8  |
|           | L x SC 2  | 4.08 x 10^5 | 2.04 x 10^5  | 0.090   | 0.914  |

The school location (L) and school category (SC) of PSSFC was selected the impactors with and analyzed by two-way ANOVA with the significance level set at 0.05. TT = total types of PSSFC; TA: total area of PSSFC; IT = types of indoor PSSFC; IA = area of indoor PSSFC; OT = types of outdoor PSSFC; OA = area of outdoor PSSFC; L = school location; SC = school category; L x SC = location and school category; SS = sum of squares; df = degree of freedom; and MS = mean squares.

https://doi.org/10.1371/journal.pone.0267009.t004

however, the location index significantly related to the scale of outdoor facilities in OA ($F = 36.143, p<0.01$) and Average OA ($F = 39.285, p<0.01$). The majority of PSSFC indicators in the Type-3 schools except IT ($p>0.05$) was significantly associated with the school location, especially in TA ($F = 100.465, p<0.01$), Average TA ($F = 64.850, p<0.01$), OT ($F = 58.283, p<0.01$) and OA ($F = 105.085, p<0.01$). In conclusion, the suburban schools possessed better PSSFC than the urban in the aspect of scale.

The school category did not significantly relate with the Type-1 schools but was yet another crucial affecting factor of PSSFC in Type-2 and Type-3 schools. More precisely, all of the PSSFC variables ($p<0.05$) in the Type-3 schools were associated with the relating school category, which determined the indoor indicators, IT ($F = 11.614, p<0.01$), IA ($F = 40.806, p<0.01$), and IA Average ($F = 16.765, p<0.01$) of indoor PSSFC in these schools. The scale of outdoor PSSFC, in the Type-2 schools, OA ($F = 14.409, p<0.01$) and Average OA ($F = 13.444, p<0.01$), had a significant link with the school category. However, when it referred to Type-1, the school category had no significant impact on those PSSFC ($p>0.05$).
Considering the interaction of school location and school category, these two factors had major impacts on the TA ($F = 10.085, p < 0.01$), OA ($F = 9.552, p < 0.01$), and IA ($F = 5.936, p < 0.01$) of Type-1 and the average outdoor space of Type-2 ($p < 0.05$). However, no significant impact on PSSFC in the Type-1 schools was found ($p > 0.05$).

The analytic results indicate that the location (suburban or urban) of primary school was associated with types (except Type-2) and area (both aggregated and averaged) of overall and outdoor facilities, rather than the types of indoor facilities. Furthermore, the suburban schools (especially in those covering more grades) were resourced with the larger space and more multiplicity of PSSFC, to accommodate more schoolchildren to participate in physical activities than the urban.

**Geographical and typological distribution**

Table 5 shows the comparison of PSSFC types in quantity and percentage in respect of the spatial and phyletic distribution. Combined with the above-mentioned results, over half of the primary schools were constructed with well-conditioned sports facilities (Type-3, the suburban and urban is 56.37% and 53.32% respectively). However, the proportion of Type-1 schools in the inner city (32.72%) was higher than those in the suburban districts (22.31%) by over 10%. Another notable finding was that the PSSFC in PS occurred to be inferior to those in 9CS and 12CS, according to the percentage of those Type-1 (35.83% of the urban PS and 27.78% of the suburban).

Fig 5 demonstrates the color-separated geographical scatter diagram, to observe the PSSFC in the scope of geographical distribution and illuminate the impact of location on PSSFC. Panel (A) illustrates the distribution of schools with clustered marks within the investigated region on the administrative map, and Panel (B) to Panel (D) presents the distribution of schools at a particular type respectively. The PSSFC levels of samples were marked by red, green, and blue, which is in line with Fig 4. Moreover, the distribution of schools within 15 km of the City Hall at all levels was listed in Fig 6 to give a clearer picture of the downtown PSSFC. It could be summarized from Figs 5 and 6 that the distribution of primary schools in the researched region showed a feature of outward diffusion from a central point. The number of primary schools in the urban area is strikingly more than that of those in the suburban. In other words, the urban contained more abundant resources for primary education. Nevertheless, the distribution of PSSFC stands out in stark contrast to the aforementioned situation: the better the school conditions were, the more uniform the geographical distribution presents;

| Location       | Type-1        | Type-2        | Type-3        |
|----------------|---------------|---------------|---------------|
| Suburban (n = 408), n (%) | 91(22.31%)    | 87(21.32%)    | 230(56.37%)   |
| 12CS, n (%)    | (0%)          | (0%)          | 13(100%)      |
| 9CS, n (%)     | 6(6.74%)      | 14(15.73%)    | 69(77.53%)    |
| PS, n (%)      | 85(27.78%)    | 73(23.85%)    | 148(48.37%)   |
| Urban (n = 437), n (%) | 143(32.72%)  | 61(13.96%)    | 233(53.32%)   |
| 12CS, n (%)    | 1(6.25%)      | 1(6.25%)      | 14(87.50%)    |
| 9CS, n (%)     | 13(21.31%)    | 6(9.84%)      | 42(68.85%)    |
| PS, n (%)      | 129(35.83%)   | 54(15.00%)    | 177(49.17%)   |

12CS = 12-year school; 9CS = 9-year school; PS = 5-year school, primary school only.

https://doi.org/10.1371/journal.pone.0267009.t005
Fig 6. Geographical distribution of PSSFC with three categories in urban districts in Shanghai by zoomed in to the scope of the urban area (distance ≤15 kilometers to the Shanghai Municipal People’s Government).

https://doi.org/10.1371/journal.pone.0267009.g006

the worse the school conditions were, the more geographical distribution was concentrated in the urban area.

Discussion

Prior studies have concerned the scientific and subjective assessment of PSSFC when investigating the correlation between children’s PA and PSSFC. As a supplement, this study sought to propose an innovative and objective method for evaluating PSSFC via machine learning.
Our findings demonstrate that the unsupervised machine learning approach on this issue performed with high interpretability and visuality in differentiating PSSFC characteristics. The results of clustering PSSFC into three types comprehensively depict the interclass differences and characteristics of PSSFC. Our experiment extends the present studies on the multidimensional analysis of PSSFC methodologically.

In the study, the location, numbers of grade levels, school category, numbers, and square meters of sports facilities were considered as evaluation indicators of PSSFC, in line with previous studies [13, 17, 18, 45]. The difference was that the data used in this study were collected from not the questionnaire but the official statistics. Among the above, the space for children’s PA had a positive correlation with both the duration and intensity of physical activity [46–48]. The type of sports facilities influenced students’ activity level by gender: boys dominated outdoor sports facilities, such as the soccer pitch, with girls preferring indoor activities [49], such as dancing and gymnastics [50]. As seen in Table 3, the quantified evaluation on PSSFC was consistent with the above finding, which suggests that a primary school with PA-demand-compliant PSSFC should be configured with 2 types of sports facilities (including 1 indoor preferred) and at least 2000 m² accessible area. Besides, the indoor PSSFC performed as a significant indicator: primary schools with at least 1 indoor facility might promote children’s MVPA, even in recess time [51]. When it refers to those without indoor facilities, the types and square meters of outdoor PSSFC were the decisive indicators for the reason that outdoor sports condition of primary school had an underlying impact on schoolchildren’s physical activities and health [18, 52, 53]. The above-mentioned findings identified the effectiveness and validity of unsupervised machine learning.

Regarding the impact of location and school category, the analytical results in Table 4 indicated that PSSFC has a striking association with the individual location, especially for the aggregated and outdoor sports space. The similarity was proved in Norway [13, 14], Taiwan [45], and Canada [54]. Although the urban PSSFC was probably affected by the shortage of land resources in developed cities and seemed to be inferior to the PSSFC in suburbs, the high density of inner-city sports facilities, particularly indoor facilities, supplements the accessibility of children’s PA [55–58]. To the best of our knowledge, it is identified that the significant links between PSSFC variables and individual school categories. In the usual sense, as the most populous type among all three types, 12CS catered to 7–12 years old students and required the best PSSFC for multigrade-shared PA in recess and after school. A possible explanation for these results may be relevant regulation of primary education in mainland China [59]. It stipulates that a PA room, greater than 40 m², must be deployed in urban 9CS and 12CS, which is not the essential requirement of PS. This explains the association between indoor facilities area and school categories.

While providing new insights into the innovation of PSSFC assessment, several limitations of this study should be clarified. Private schools, which are the crucial part of primary education, were not involved. The omission of these schools may have resulted in some potential bias, although it is difficult to make assertions on the direction and size of this bias. Furthermore, although the geographic factors of urban and suburban areas were considered, the influence of geomorphology on the layout of primary school was not taken into account. The results are specific to primary schools in mainland China, and as such are likely to be context-specific. Moreover, the analysis is representative and cross-sectional in the developed region, which may not apply to the rest of world. Further research in different countries among different landforms would be beneficial. The approach of visualization and evaluation on PSSFC is exclusively based on objective data. Due to the lack of more detailed indicators in the NSFC, the following potential factors should be involved in further practical applications and studies: subjective assessment, the student population in each school, PA space per student, and land.
occupation and floor space of each school. Last but not least, constrained by the lack of data and the restriction of any experiment in primary schools induced by the COVID-19 pandemic, this study was barely limited to investigating the PSSFC without the analysis of results with the students’ PA and PE in sampled schools. Despite these limitations, this study provides a demonstration of interdisciplinary research between artificial intelligence and the PA environment and analyzes the spatial and phyletic characteristics of PSSFC in developing countries. The evaluation method on PSSFC can be used to inform the regionalized and personalized intervention for children’s PA promotion, according to the current state of individual PSSFC.

Conclusions

The finding in this paper verified that combined-approach unsupervised learning facilitates understanding the characteristics of sports facilities in primary schools in a visual way. The approach used in this study had acceptable levels of interpretability and was found to effectively differentiate sports facilities in primary schools. In addition, with the statistical analysis of the classification results, the spatial and typological association with the variables of sports facilities in primary schools was discovered and was consistent with previous studies. Our findings reveal that the regional school sports facility condition in primary schools has geographical distribution characteristics. Primary schools with more grades of students are equipped with more types and sizes of sports facilities.

Further work is aimed to quantify the PSSFC association with more varied factors, such as subjective remarks, attending numbers of students, and other potential influences.

Supporting information

S1 Fig. Comparison of classified overall PSSFC by school type and location. (TIF)

S2 Fig. Comparison of classified indoor PSSFC by school type and location. (TIF)

S3 Fig. Comparison of classified outdoor PSSFC by school type and location. (TIF)

S1 Dataset. Data of sampled PSSFC. (PDF)

S1 File. (CSV)

Author Contributions

Conceptualization: Jun Xia.

Data curation: Jun Xia, Hua Chen.

Formal analysis: Jun Xia.

Methodology: Jie Zhuang, Zhenbo Cao.

Project administration: Zhenbo Cao.

Software: Jun Xia.

Supervision: Jie Zhuang, Peijie Chen.

Validation: Jun Xia, Hua Chen.
Visualization: Jun Xia.

Writing – original draft: Jun Xia.

Writing – review & editing: Jihong Wang, Jie Zhuang, Zhenbo Cao, Peijie Chen.

References

1. Kelly IR, Phillips MA, Revels M. Contribution of the school environment to physical fitness in children and youth. Journal of Physical Activity and Health. 2010; 7:333–342. https://doi.org/10.1123/jpah.7.3.333 PMID: 20551489

2. World Health Organization. Global recommendations on physical activity for health. Geneva: World Health Organization, 2010.

3. Troiano RP, Flegal KM. Overweight prevalence among youth in the United States: why so many different numbers? Int J Obes Relat Metab Disord 1999; 23:S22–S27. https://doi.org/10.1038/sj.ijo.0800855 PMID: 10340801

4. Ogden CL, Flegal KM, Carroll MD, Johnson CL. Prevalence and trends in overweight among U.S. children and adolescents, 1999–2000. JAMA 2002; 288:1728–1732. https://doi.org/10.1001/jama.288.14.1728

5. Gordon-Larsen PL, Adair SL, Nelson MC. Five-year obesity incidence in the transition period between adolescence and adulthood: the National Longitudinal Study of Adolescent Health. Am J Clin Nutr 2004; 80:569–575. https://doi.org/10.1093/ajcn/80.3.569 PMID: 15321794

6. Slyper A. The pediatric obesity epidemic causes and controversies. J Clin Endocrinol Metab. 2004; 89:2540–2547. https://doi.org/10.1210/jc.2003-031449 PMID: 15181021

7. Biddle SJ, Gorely T, Stensel DJ. Health-enhancing physical activity and sedentary behavior in children and adolescents. J Sports Sci. 2004; 22:679–701. https://doi.org/10.1080/02640410410001712412 PMID: 15370482

8. Bauman AE, Reis RS, Sallis JF. Correlates of physical activity: Why are some people physically active and others not? Lancet. 2012; 380, 258–271. https://doi.org/10.1016/S0140-6736(12)60735-1 PMID: 22818938

9. Robertson-Wilson JE, Leatherdale ST, Wong SL. Social-ecological correlates of active commuting to school among high school students. J Adolesc Health. 2008; 42, (5):486–95. https://doi.org/10.1016/j.jadohealth.2007.10.006 PMID: 18407044

10. O’Malley PM, Johnston LD, Delva J, Terry-McElrath YM. School physical activity environment related to student obesity and activity: a national study of schools and students. J Adolesc Health. 2009; 45 (3Suppl): S71–81. https://doi.org/10.1016/j.jadohealth.2009.04.006 PMID: 19694440

11. McGrath LJ, Hopkins WG., Hinckson EA. Associations of objectively measured built-environment attributes with youth moderate-vigorous physical activity: a systematic review and meta-analysis. Sports Medicine. 2015; 45(6):841–865. https://doi.org/10.1007/s40279-015-0301-3 PMID: 25618013

12. Sallis JF, Frank LD, Pate RR. Correlates of physical activity: Why are some people physically active and others not? JAMA. 2000; 283:136–147. https://doi.org/10.1001/jama.283.2.136 PMID: 10709180

13. Huag E, Torsheim T, Samdal O. Physical environmental characteristics and individual interests as correlates of physical activity in Norwegian secondary schools: the health behavior in school-aged children study. Int J Behav Nutr Phys Act. 2008; 5(1):47–56. https://doi.org/10.1186/1479-5868-5-47 PMID: 18823645

14. Huag E, Torsheim T, Samdal O. Local school policies increase physical activity in Norwegian secondary schools. Health Promot Int. 2009; 24(1):63–72. https://doi.org/10.1093/heapro/dap040 PMID: 19984244

15. Ishii K, Shibata A, Sato M, Oka K. Int J Environ Res Public Health 2014; 11:7195–7206. https://doi.org/10.3390/ijerph110707195 PMID: 25029495

16. Rezende LF, Azeredo CM, Silva KS; Claro RM. The role of school environment and after-school physical activity with health-related physical fitness among junior high school students in Taiwan. Int J Environ Res Public Health 2017; 14(1): E83. https://doi.org/10.3390/ijerph14010083 PMID: 28098836

17. Huag E, Torsheim T, Samdal O. Title of Thesis. The characteristics of the outdoor school environment associated with physical activity. Health Educ Res. 2010, 25(2):248–256. https://doi.org/10.1093/her/cyn050 PMID: 19936270
19. Bevans BK, Fitzpatrick LA, Sanchez BM. Physical education resources, class management, and student physical activity levels: a structure-process-outcome approach to evaluating physical. J School Health 2010; 80(12):573–580. https://doi.org/10.1111/j.1746-1561.2010.00544.x PMID: 21087253

20. The Statistics Bureau of Shanghai. Shanghai statistical yearbook 2018. Beijing: China Statistic Press; 2019, pp. 154.

21. Chopde NR, Nichat MK. Landmark based shortest path detection by using A* and Haversine formula. Int J Innovative Res Comput Commun Eng. 2013; 1(2):298–302.

22. Nayak J, Naik B, Behera HS. Fuzzy c-means(FCM) clustering algorithm: a decade review from 2000 to 2014. (2015) Proceedings of the International Conference on CIDM. New Delhi: Springer; 2015. (2):133–149. https://doi.org/10.1007/978-81-322-2208-8_14

23. Celebi ME, Kingravi HA, Vela PA. A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Syst Appl 2013; 40(1):200–210. https://doi.org/10.1016/j.eswa.2012.07.021

24. Heil J, Häring V, Marshner B, et al. Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance leaf spectra: a case study with west African soils. Geoderma 2019; 337:11–12. https://doi.org/10.1016/j.geoderma.2018.09.004

25. Bezdek JC, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. Comput Geosci 1984; 191–203. https://doi.org/10.1016/0098-3004(84)90020-7

26. Choudhry MS, Kapoor R. Performance Analysis of fuzzy c-means clustering methods for MRI image segmentation. Procedia Computer Science 2016; 89: 749–758. https://doi.org/10.1016/j.procs.2016.06.052

27. Zeng Z, Han C, Wang L, et al. Unsupervised brain tissue segmentation by using bias correction fuzzy c-means and class-adaptive hidden Markov random field modeling. Lect Notes Electr Eng 2014; 269:579–587 https://doi.org/10.1007/978-94-007-7618-0_56

28. Lai DTC, Garibaldi JM. A preliminary study on automatic breast cancer data classification using semi-supervised fuzzy c-means. Int J Biomed Eng Technol 2013; 13(4):303–322. https://doi.org/10.1504/IJBET.2013.058355

29. Phu VN, Dat ND, Tran N. et al. Fuzzy C-means for English sentiment classification in a distributed system. Appl Intell 2017; 46:717–738. https://doi.org/10.1007/s10489-016-0858-z

30. Mahela OP, Shaik AG. Power quality improvement in distribution network using DSTATCOM with battery energy storage system. Int J Elec Power 2016; 83: 229–240. https://doi.org/10.1016/j.ijepes.2016.04.011

31. Guo M, Yang N. Features-clustering-based earth fault detection using singular-value decomposition and fuzzy c-means in resonant grounding distribution systems. Int J Elec Power 2017; 93:97–108. https://doi.org/10.1016/j.ijepes.2017.05.014

32. Haldar NA, Khan FA, Ali A. Arrhythmia classification using Mahalanobis distance based improved Fuzzy C-Means clustering for mobile health monitoring systems. Neurocomputing 2017; 220:221–235. https://doi.org/10.1016/j.neucom.2016.08.042

33. Anter AM, Ali M. Feature selection strategy based on hybrid crow search optimization algorithm integrated with chaos theory and fuzzy c-means algorithm for medical diagnosis problems. Soft Comp 2020; 24(3):156–1584. https://doi.org/10.1007/s00500-019-03988-3

34. Caron M, Bojanowski P, Joulin A, Douze M. Deep clustering for unsupervised learning of visual features. Proceeding of the European conference on computer vision, Munich, Germany. Berlin: Springer; 2018, pp:139–156. https://doi.org/10.1007/978-3-030-01264-9_9

35. Bhatia V, Rani R. DFuzzy: a deep learning-based fuzzy clustering model for large graphs. Knowledge and Information Systems. 2018; 57(8):159–181. https://doi.org/10.1007/s10115-018-1156-3

36. Izakian H, Abraham A. Fuzzy c-means and fuzzy swarm for fuzzy clustering problem. Expert Syst Appl. 2011; 38(3):1835–1838. https://doi.org/10.1016/j.eswa.2010.07.112

37. Wu K, Yang M. A cluster validity index for fuzzy clustering. Pattern Recogn Lett. 2005; 26(9):1275–1291. https://doi.org/10.1016/j.patcog.2007.10.004

38. van der Maaten L, Hinton G. Visualizing high-dimensional data using t-SNE. J Mach Learn 2008; 9:2579–2605.

39. van der Maaten L. Accelerating t-SNE using tree-based algorithms. J Mach Learn 2014; 15:3221–45.

40. Mahfouz A, van der Giessen M, van der Maaten L, et al. Visualizing the spatial gene expression organization in the brain through non-linear similarity embeddings. Methods 2015; 73: 79–89. https://doi.org/10.1016/j.ymer.2014.10.004 PMID: 25449901
41. Schelker M, Feau S, Du J, et al. Estimation of immune cell content in tumor tissue using single-cell RNA-seq data. Nat Commun 2017; 8: 2032. https://doi.org/10.1038/s41467-017-02289-3 PMID: 29230012

42. Oliveira HM, Zaruz TF, Vieira MF, Andrade AO. t-SNE applied to discriminate healthy individuals from those with Parkinson’s disease executing motor tasks detected by non-contact capacitive sensors. IFMBE Proceedings World Congress on Medical Physics and Biomedical Engineering 2018; 68:1. https://doi.org/10.1007/978-981-10-9035-6_78

43. Song W, Wang L, Liu P, et al. Improved t-SNE based manifold dimensional reduction for remote sensing data processing. Multimed Tools Appl 2019; 78, 4311–4326. https://doi.org/10.1007/s11450-018-5715-0

44. Chen J, Wang C, Wang B, et al. A visualized classification method via t-distributed stochastic neighbor embedding and various diagnostic parameters for planetary gearbox fault identification from raw mechanical data. Sensors and Actuators A: Physical 2018; 284:52–65. https://doi.org/10.1016/j.sna.2018.10.021

45. Chen J, Unnithan VB, Kennedy C. Correlates of physical fitness and activity in Taiwanese children. Int Nurs Rev. 2008; 55(1):81–89. https://doi.org/10.1111/j.1466-7657.2007.00588.x PMID: 1827540

46. Thompson JL, Davis SM, Gittelsohn J. Patterns of physical activity among American Indian children: an assessment of barriers and support. J Community Health. 2001; 26(6): 423–445. https://doi.org/10.1023/a:1012507323784 PMID: 11759094

47. Beighle A, Morgan CF, Le Masurier G, Pangrazi RP. Children’s physical activity during recess and outside of school. J Sch Health. 2006; 76:516–520. https://doi.org/10.1111/j.1746-1561.2006.00151.x PMID: 17096825

48. Parrish A, Yeatman H, Iverson D. Using interviews and peer pairs to better understand how school environments affect young children’s playground physical activity levels: a qualitative study. Health Educ Res. 2012; 27(2):269–280. https://doi.org/10.1093/her/cyr049 PMID: 21712499

49. Ng W, Bhat H, Ntoumanis N, et al. Physical activity across ages and seasons in Swedish second, fifth, and eighth-graders. BMC Public Health. 2014; 14:803. https://doi.org/10.1186/1471-2458-14-803 PMID: 25099142

50. Pagels P, Raustorp A, De Leon AP, Mårtensson F, Kylin M, Boldemann C. A repeated measurement study investigating the impact of school outdoor environment upon physical activity across ages and seasons in Swedish second, fifth, and eighth-graders. BMC Public Health. 2014; 14:803. https://doi.org/10.1186/1471-2458-14-803 PMID: 25099142

51. Haug E, Torsheim T, Samdal O. Physical environmental characteristics and individual interests as correlates of physical activity in Norwegian secondary schools: the health behavior in school-aged children study. Int J Behav Nutr Phys Act. 2008; 5:47. https://doi.org/10.1186/1479-5868-5:47 PMID: 18828545

52. Hobin E, Leatherdale ST, Manske S. Are environmental influences on physical activity distinct for urban, suburban, and rural schools? A multilevel study among secondary school students in Ontario, Canada. J School Health 2013; 83(5):357–367. https://doi.org/10.1111/josh.12039 PMID: 23517004

53. Huang J, Shen GQ, Zheng HW. Is insufficient land supply the root cause of housing shortage? Empirical evidence from Hong Kong. Habitat Int. 2015; 49:538–546. https://doi.org/10.1016/j.habitatint.2015.07.006

54. Wong BY, Cerin E, Ho SY. Adolescents’ physical activity: competition between perceived neighborhood sports facilities and home media resources. Int. J. Pediatr Obes. 2010; 5(2):169–176. https://doi.org/10.3109/1747716903159432 PMID: 19657859

55. Prins RG, Ball K, Timperio A, Salmon J, Onemaac A, Brug J, et al. Associations between the availability of facilities within three different neighborhood buffer sizes and objectively assessed physical activity in adolescents. Health Place. 2011; 17(6): 1228–1234. https://doi.org/10.1016/j.healthplace.2011.07.012 PMID: 21889390

56. Reimers AK, Wagner M, Alvanides S. Proximity to sports facilities and sports participation for adolescents in Germany. PloS ONE 2014; 9(3):e93059. https://doi.org/10.1371/journal.pone.0093059 PMID: 24675689

57. Standards for construction of school buildings for urban primary and secondary schools. National center for schooling development program. Available online: http://www.cspd.edu.cn/article/589.html.