Penerbit Akademia Baru

Journal of Advanced Research in Computing and Applications

Journal homepage:
https://akademiabaru.com/submit/index.php/arca/index
ISSN: 2462-1927

Deep-Learning Pre-Processing for Improvement Of K- Means Cluster Analysis of Seniors' Walkability in Hradec Kralove And Ostrava (Two Middle-Sized Czech Cities)

Ondrej Krejcar[1], Pavel Kukuliac [2], Lim Kok Cheng[3], Ali Selamat[4,*], Jiri Horak[5]

[1] Center for Basic and Applied Research Faculty of Informatics and Management, University of Hradec Kralove Hradec Kralove, Czech Republic
[2] Department of Geoinformatics Faculty of Mining and Geology Technical University of Ostrava, Ostrava, Czech Republic
[3] Department of Computing Universiti Tenaga Nasional, Selangor, Malaysia
[4] Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) Kuala Lumpur, Malaysia
[5] Department of Geoinformatics Faculty of Mining and Geology Technical University of Ostravac, Ostrava, Czech Republic

ARTICLE INFO

Article history:
Received 10 July 2022
Received in revised form 19 August 2022
Accepted 3 September 2022
Available online 20 September 2022

Keywords:
Cluster analysis, walkability, elderly, deep learning, k-means, HDBscan, N2D

ABSTRACT

Increasing people's well-being, reducing traffic, and creating a healthy urban environment in contemporary cities is quite dependent on excellent conditions for walking and generally for supporting human physical activity. Various indicators and metrics exist to assess walking conditions. One of them is the walking index (WAI), representing indicators based on the environment. Walkability receives growing interest for vulnerable group of people, namely higher age elderly. Due to erosion of physical and financial capabilities, walking is considered as a safe and saving form of physical exercise to maintain health status, a platform for positive direct and indirect effects – improvement of quality of life (QoL), health status, well-being; decreasing of traffic and related pollution; direct economic effect (quotations). Our motivation was to improve classic cluster analysis of urban walking conditions for the elderly to obtain more specific and robust classification, assess the potential of advanced machine-learning based clustering methods to discover more specific classes of urban conditions to better address improvement of urban conditions using specific urban planning measures. And because classic K-means does not provide satisfactory results, we focused also on HDBScan, Soft Clustering and N2D method. Finally, the results proved the N2D method is the efficient method of clustering and provides improved results for urban walkability characteristics.

1. Introduction

Today's cities focus on their inhabitants and aim to improve the urban environment through a complex set of activities. "Pedestrian first" is a key term for many city representatives. Such improvements positively affect most seniors whether or not they themselves feel physically handicapped. Due to the gradual decline of physical capabilities and frequent temporal restrictions due to episodic health issues in seniors, any decrease in mobility barriers is welcome [1].

* *Corresponding author.*
E-mail address: aselamat@utm.my (Ali Selamat)

Various indicators and walking conditions. One of them is the walking index (WAI), representing indicators based on the environment.

Cluster analysis [2-4] is a group of frequent multivariate statistical methods, different approaches, issues. Cluster analysis or clustering is also the task of grouping a set of objects (clusters) which are more like each other than to those in other groups (clusters). Classic k-means [5] as one of the possible solutions in our case does not provide satisfactory results, thus we tested several other cluster methods.

Motivation was to improve classic cluster analysis of urban walking conditions for elderly to obtain more specific and robust classification, assess the potential (capacity, performance) of advanced machine-learning based clustering methods to discover more specific classes of urban conditions to better address improvement of urban conditions using specific urban planning measures [6-8].

2. Cluster Analysis

Clustering can be formulated as a multi-objective optimization problem [9–12]. The appropriate clustering algorithm and parameter settings depend on the individual data set and intended use of the results. For our case we taken into consideration the following used methods:

K-means – k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centres or cluster centroid), serving as a prototype of the cluster [5].

HDBscan – is a hierarchical version of DBSCAN which is also faster than OPTICS, from which a flat partition consisting of the most prominent clusters can be extracted from the hierarchy [13].

Soft clustering – means that each object belongs to each cluster to a certain degree (for example, a likelihood of belonging to the cluster) [3]. Results are accompanied with probability of the resulting classification.

N2D method (N2D) – that effectively replaces the clustering network with a manifold learning technique on top of the autoencoded representation [14]. Three main steps are as following:

(i) Create an autoencoder (Deep learning part) that will learn lower dimensional representation of the data, which will hopefully capture the most important information and structures within it.

(ii) Apply a manifold learning method (such as UMAP or TSNE) to further reduce the dimensions of the data and create finer representations that will improve the performance of the clustering algorithm.

(iii) Cluster the data (K-means, HDBSCAN, etc.)

For evaluation of cluster analysis results various methods and metrics are applied.

HDBSCAN monitors the DBCV score (Density Based Cluster Validity) indicator. The closer the score value is to 1, the more plausible the clusters.

3. Walkability (Characteristics)

Current cities aim to improve living conditions for their inhabitants and visitors. One of the important trends is to improve walking conditions which have a platform positive direct and indirect effects – improvement of quality of life (QoL), health status, well-being; decreasing of traffic and related pollution; direct economic effect (quotations) [15-18].

Walkability receives growing interest for vulnerable group of people, namely higher age elderly. Due to erosion of physical and financial capabilities walking is considered as a safety and saving form of physical exercises to maintain health status. Walking for elderly is also important to keep social contacts. Furthermore, seniors above 85 ages are usually limited to drive and the walking accessibility become an important factor to keep their personal self-reliance and independency (quotations) [1], [19].

To measure walking condition different characteristics and indicators are used. Among objective-based indicators is the Walkability Index (WAI) developed by Frank *et al.,* [1]. WAI evaluates connectivity, heterogeneity of land use, shopping area, and household density.

WAI aggregates four indicators: the connectivity index (CONN), which measures the density of intersections of walkable roads, Shannon's entropy index (ENT), which quantifies the heterogeneity of land uses within an area, the floor area ratio (FAR) index, which evaluates the intensity of shopping opportunities as a ratio of floor area and available commercial land use, and the household density index (HDENS), which is related to residential land use.

(Case study) Walkability is assessed in two middle sized Czech cities:

- Hradec Kralove (population 100,000) is a typical old central European city with an old city core, and which was developed mainly after the decline of fortification in the 19th century [20].
- Ostrava (population 290,000) consists of an agglomeration of relatively closed communities (urban blocks) separated by crop fields, forests, and industrial parks because of short but intensive industrial development and administrative union of originally independent municipalities.

There are also input data defined as following:

- CROSS, ENT, FAR, HDENS, - see above
- DAMENITY – density of amenities. Amenity - weighted average of 4 main categories of amenities important for elderly: (a*RETAIL+ b*DHMARKET+ c*DCULTURE + d*DDOCTOR + e*DPARK) with weights: a = 4/sum, b=2/sum, c=1/sum, d=1/sum, e=1/sum according to WalkScore weighing system (WalkScore methodology)
- CONPATH – density of pathways + service roads
- BLOCK – average distance between crossings
- SLOPEMEAN – average slope on pathways

Next step in processing we made was Z-score standardization [1]. Spatial distribution can be seen in the figure (Figure 1).

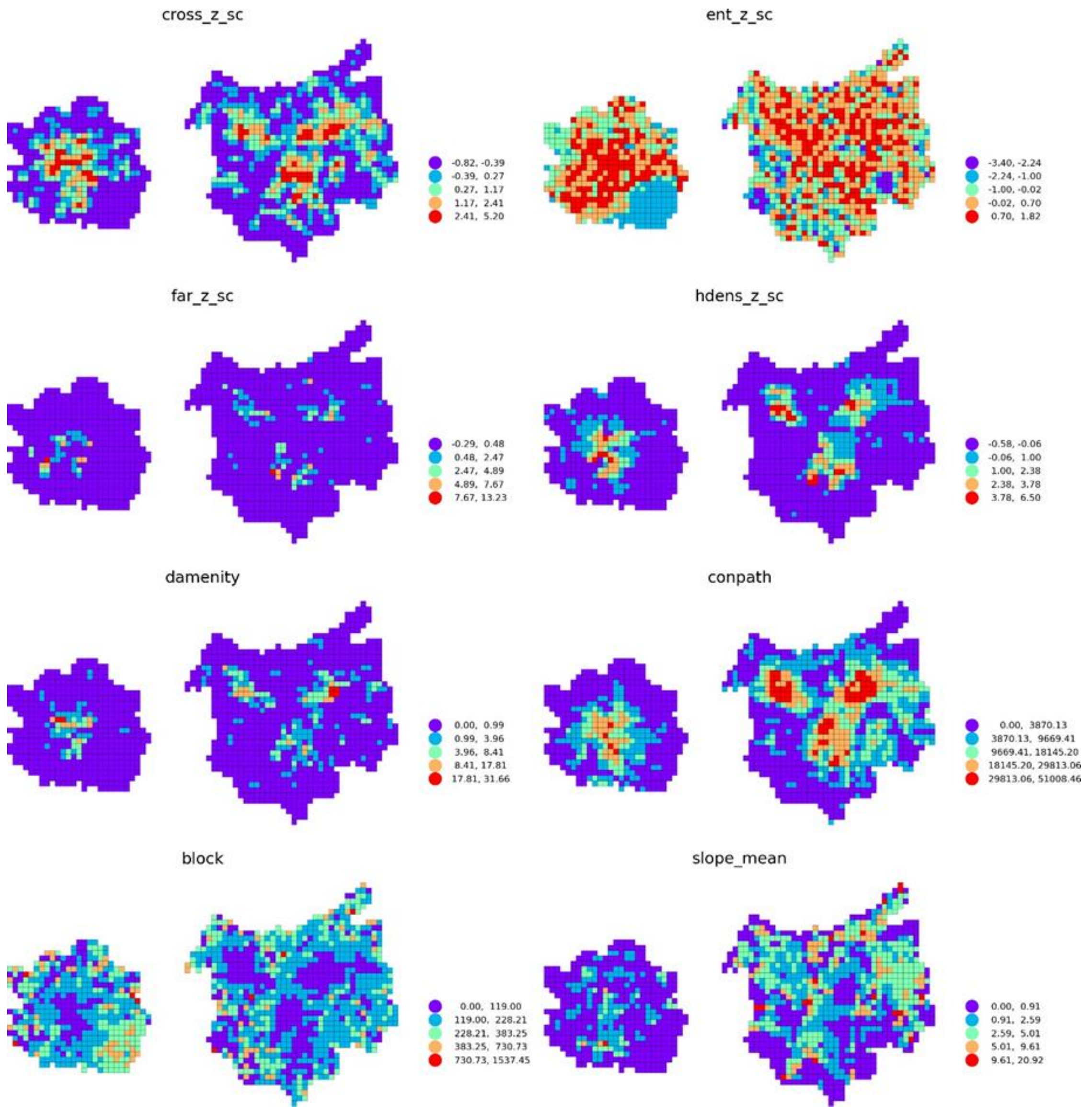

Fig. 1. Spatial distribution of input environmental walking characteristics (z-score) in Hradec Kralove and Ostrava

4. Discussion and Results

The outputs are stored in separate folders according to the type of algorithm. The final data (shp and excel) are available in the Data_results folder, where the relevant identifiers of individual clusters according to the type of algorithm are listed for each grid cell. The HDBSCAN algorithm also generates a cluster membership probability (cluster_prob) [13].

There are always 3 outputs:

- Full analysis using autoencoder + manifold learning + clustering algorithm -> best results
- Using only manifold learning + clustering algorithm without using deep learning -> decent results
- Using only the clustering algorithm without pre- processing by reducing the dimensionality of the data -> bad results

Journal of Advanced Research in Computing and Applications
Volume 28, Issue 1 (2022) 1-11

HDBscan results are visible at Figure 2-7.

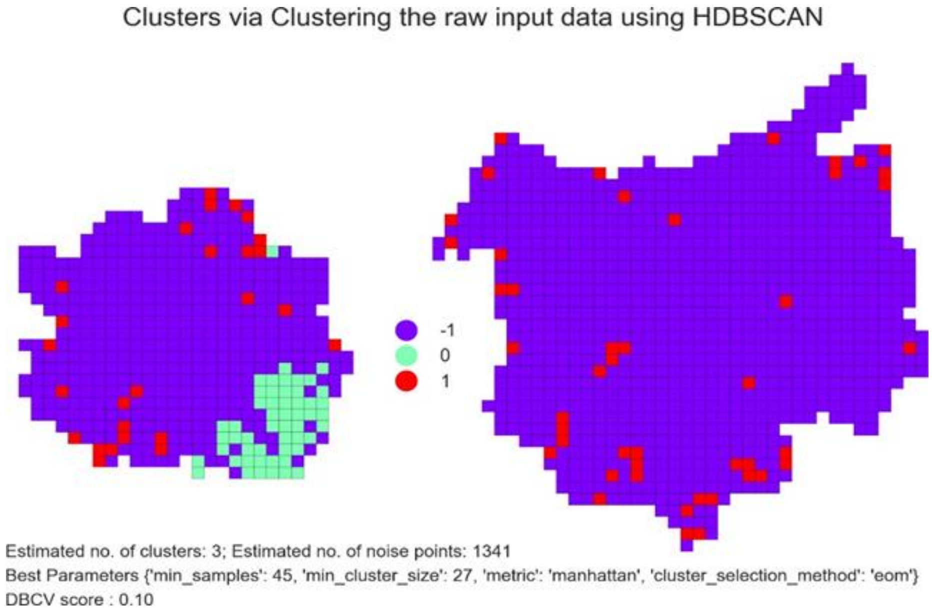

Fig. 2. Clustering HDBscan with original data

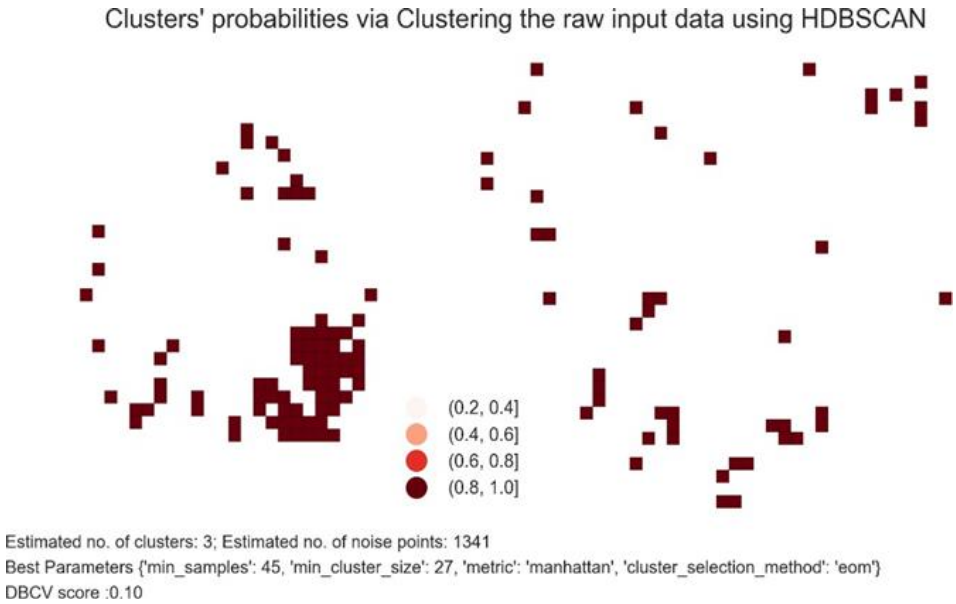

Fig. 3. Clustering HDBscan with original data – probabilities

Journal of Advanced Research in Computing and Applications
Volume 28, Issue 1 (2022) 1-11

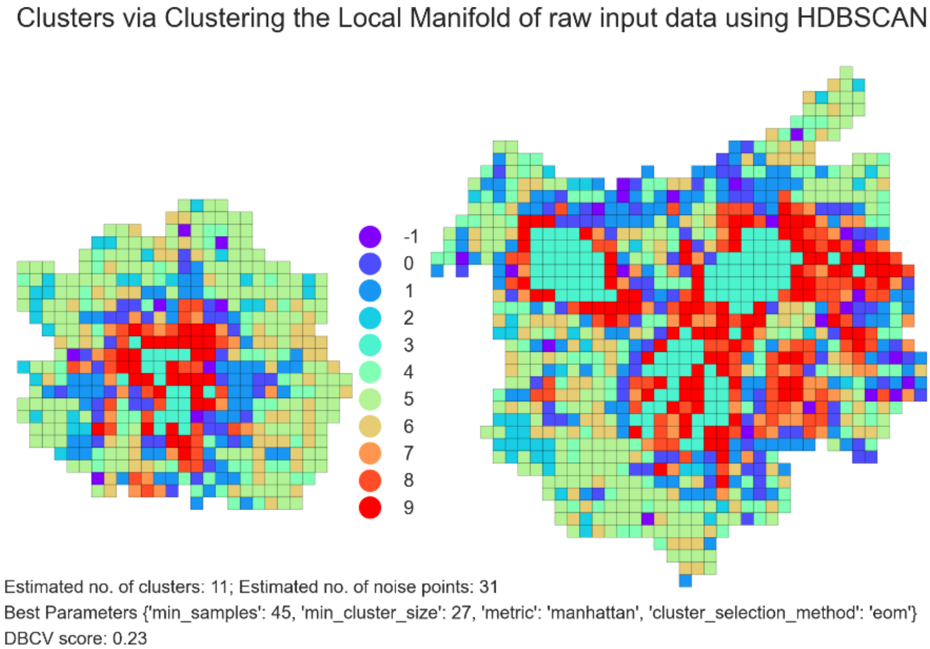

Fig. 4. Clustering HDBscan on manifold learning

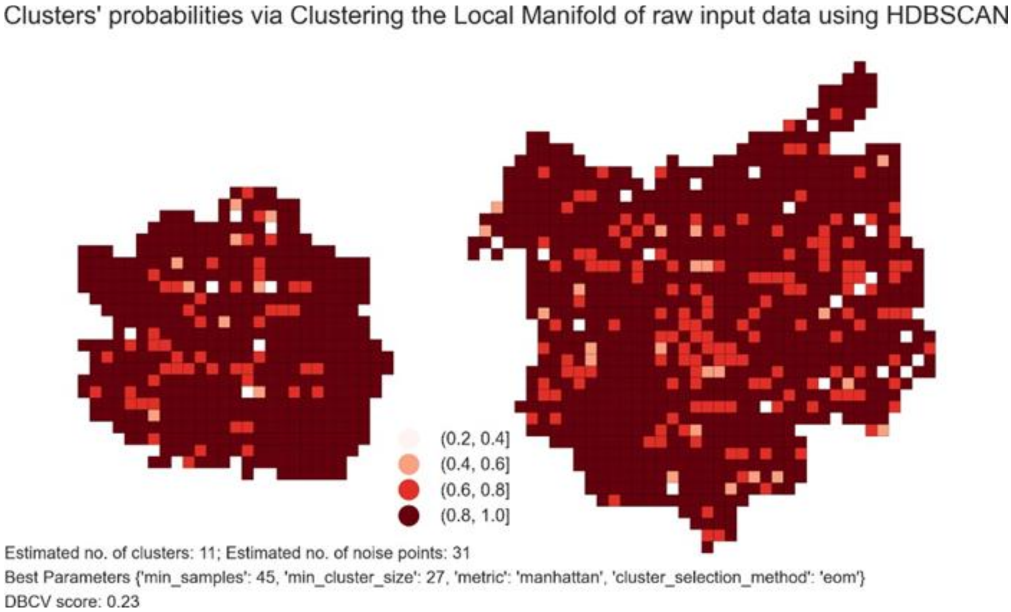

Fig. 5. Clustering HDBscan on manifold learning - probabilities

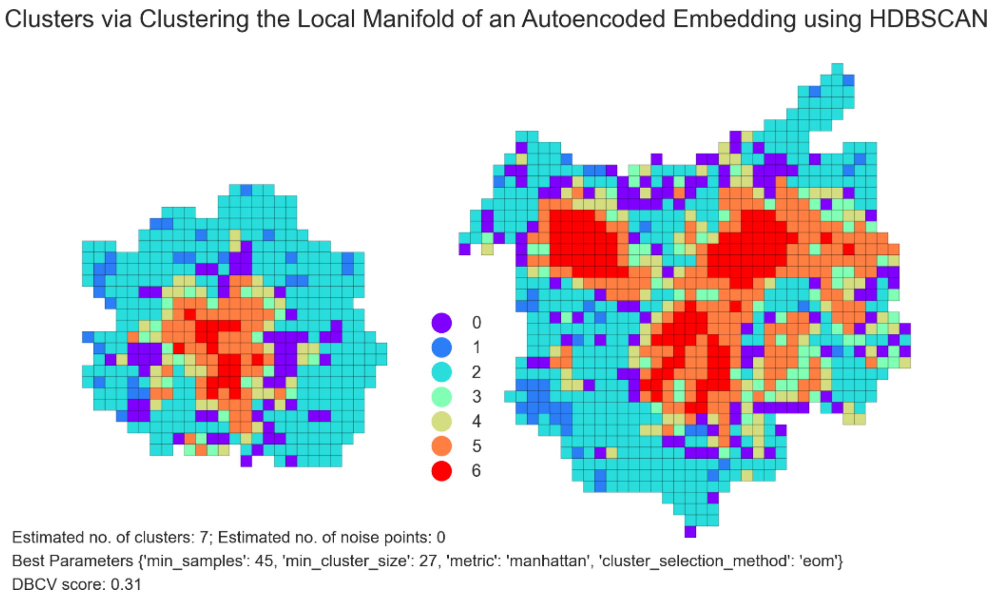

Fig. 6. Clustering HDBscan on manifold learning + deep learning (full N2D method)

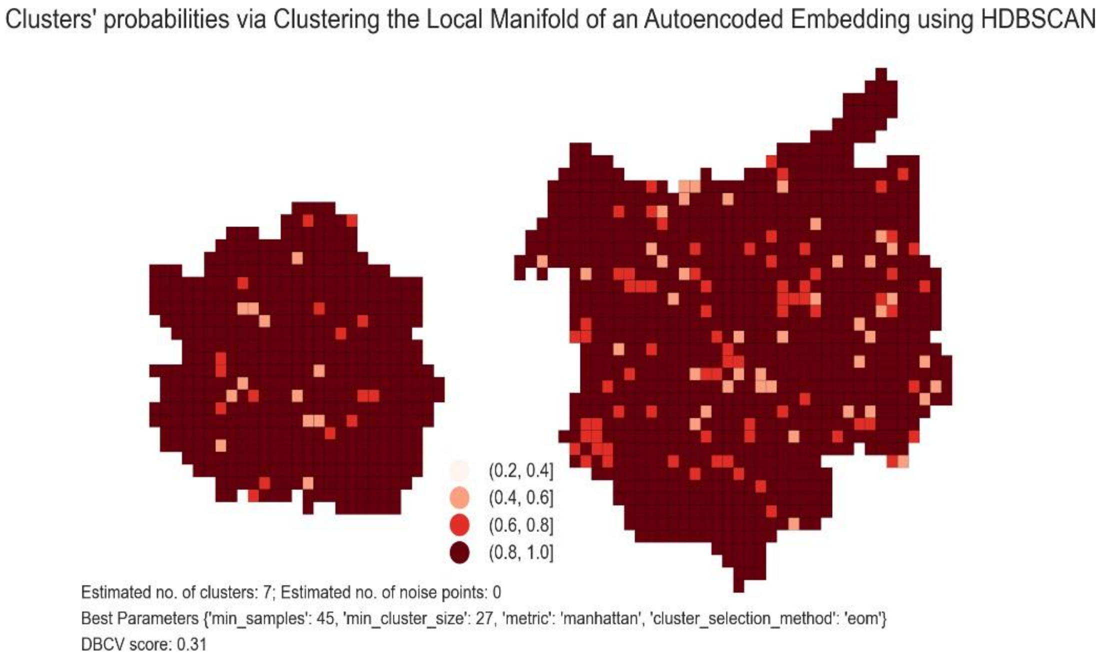

Fig. 7. Clustering HDBscan on manifold learning + deep learning (full N2D method)-probabilities

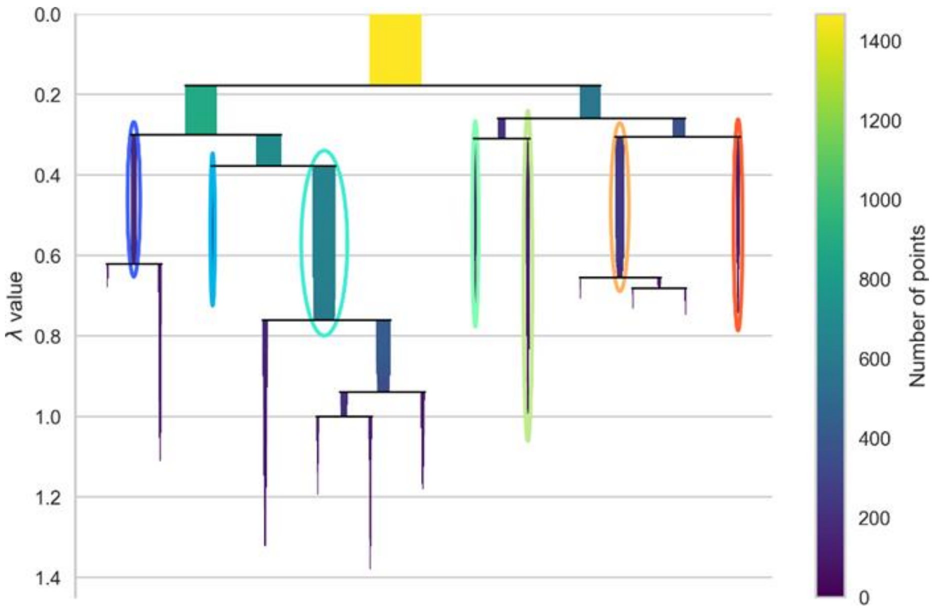

Fig. 8. Condensed hierarchical clustering tree for N2D output

The similar overview can be seen for K-means at the figures below (Figure 9-11).

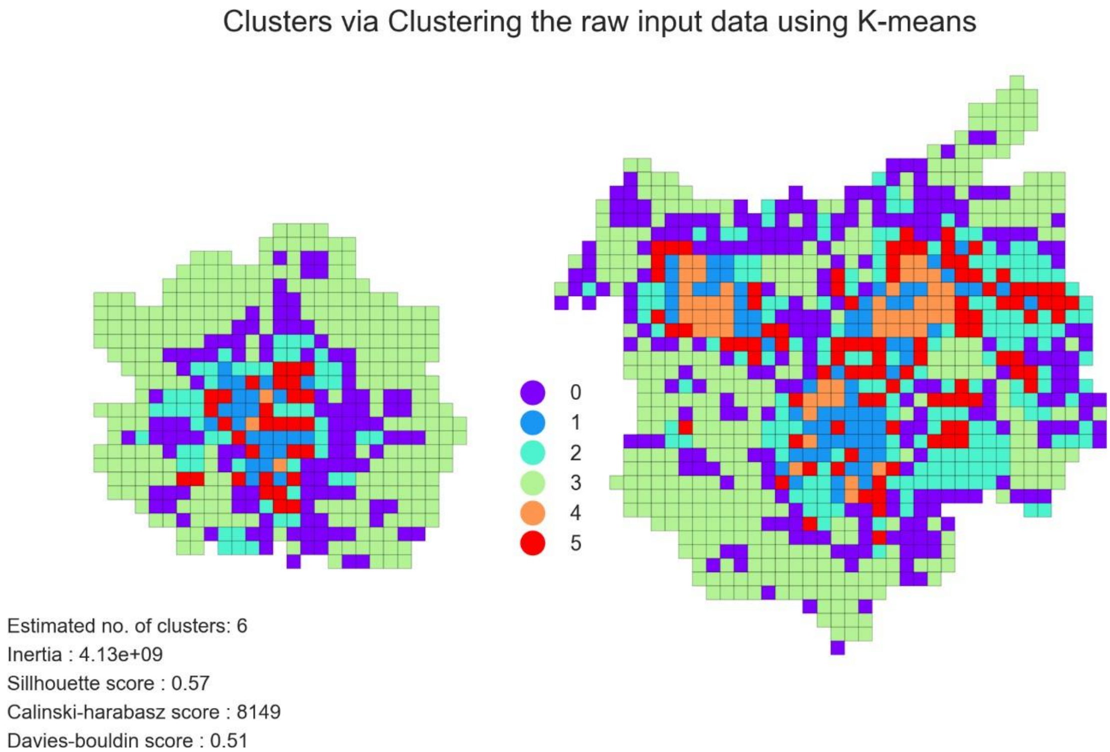

Fig. 9. Cluster via Clustering input data using K-means

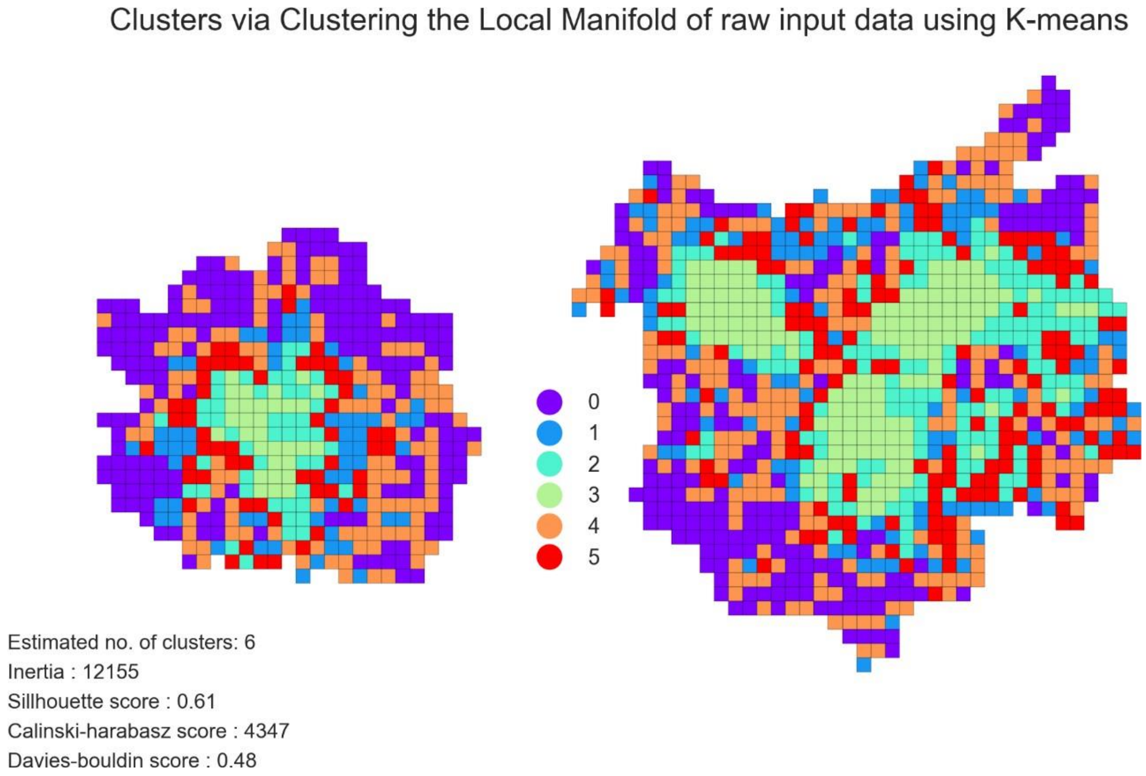

Fig. 10. Cluster via Clustering the Local Manifold of raw input data using K-means

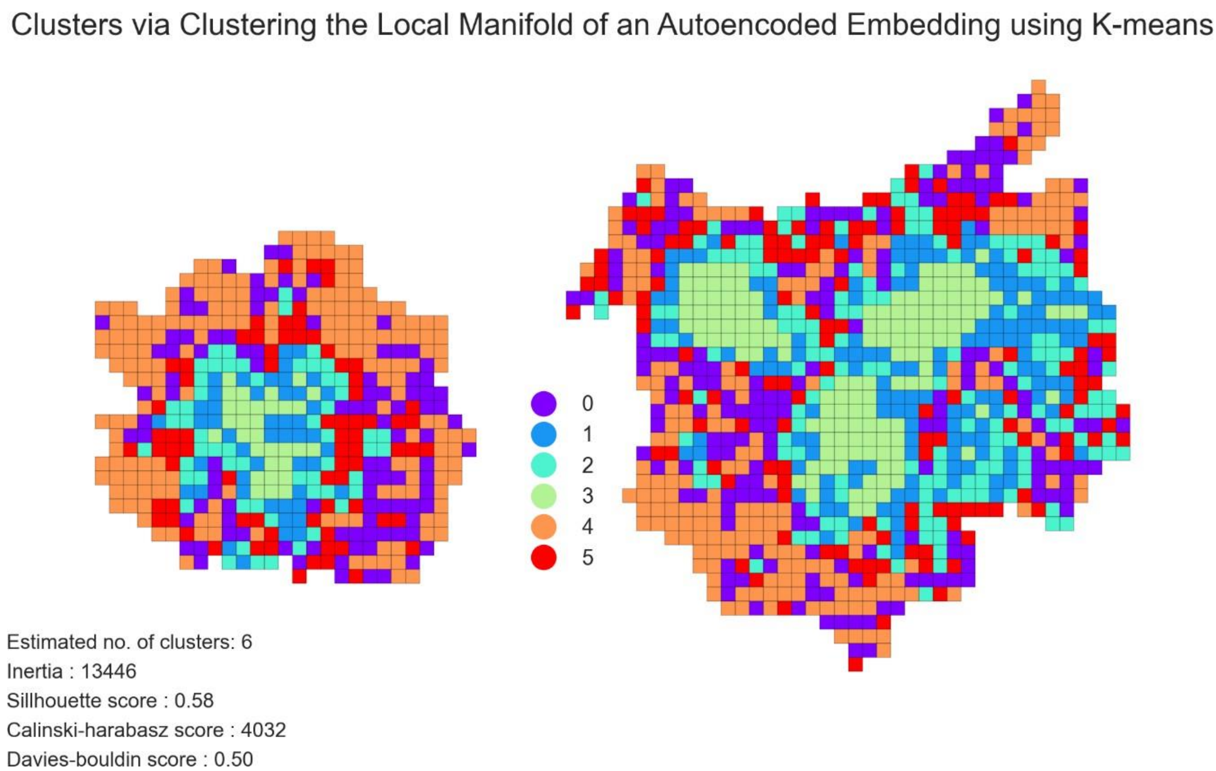

Fig. 11. Cluster via Clustering the Local Manifold of an Autoencoded Embedding using K-means

5. Conclusions

Basic motivation for this paper was to improve classic cluster analysis of urban walking conditions for elderly to obtain more specific and robust classification. Because the classic K-means don't provide satisfactory results, we focussed also on HDBScan, Soft Clustering and N2D method. Finally, it was proved the N2D method is the efficient method of clustering and provides improved results for urban walkability characteristics.

Acknowledgment
This research was funded by the Czech Science Foundation, grant number 21-22276S ''Accessibility modelling for seniors, perception of accessibility and determinants of their spatial mobility".

References
[1] Horak, Jiri, Pavel Kukuliac, Petra Maresova, Lucie Orlikova, and Ondrej Kolodziej. "Spatial pattern of the walkability index, walk score and walk score modification for elderly." ISPRS International Journal of Geo-Information 11, no. 5 (2022): 279. https://doi.org/10.3390/ijgi11050279

[2] Ivanenko, L. V., E. A. Karaseva, and E. P. Solodova. "Clusters, digital economy and smart city." In *Digital Transformation of the Economy: Challenges, Trends and New Opportunities*, pp. 291-295. Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-11367-4_28

[3] Cheng, Lim Kok, Ali Selamat, Mohd Hazli Mohamed Zabil, Md Hafiz Selamat, Rose Alinda Alias, Fatimah Puteh, Farhan Mohamed, and Ondrej Krejcar. "Usability prioritization using performance metrics and hierarchical agglomerative clustering in MAR-learning application." In *New Trends in Intelligent Software Methodologies, Tools and Techniques*, pp. 731-744. Ios Press, 2017.

[4] Sekiguchi, Takuya, Kohei Tamura, and Naoki Masuda. "Population changes in residential clusters in Japan." *Plos one* 13, no. 5 (2018): e0197144. https://doi.org/10.1371/journal.pone.0197144

[5] Zhao, Shizhen, Wenfeng Li, and Jingjing Cao. "A user-adaptive algorithm for activity recognition based on k-means clustering, local outlier factor, and multivariate gaussian distribution." *Sensors* 18, no. 6 (2018): 1850. https://doi.org/10.3390/s18061850

[6] Arranz-López, Aldo, Luis M. Mejía-Macias, and Julio A. Soria-Lara. "Combining walking accessiblity measures to map spatial inequalities." *Journal of Maps* 17, no. 1 (2021): 85-94. https://doi.org/10.1080/17445647.2021.1962752

[7] Brüchert, Tanja, Pia Hasselder, Paula Quentin, and Gabriele Bolte. "Walking for transport among older adults: a cross-sectional study on the role of the built environment in less densely populated areas in northern Germany." *International journal of environmental research and public health* 17, no. 24 (2020): 9479. https://doi.org/10.3390/ijerph17249479

[8] Hou, C., Z. Ping, K. Yang, S. Chen, X. Liu, H. Li, M. Liu et al. "Trends of activities of daily living disability situation and association with chronic conditions among elderly aged 80 years and over in China." *The Journal of nutrition, health and aging* 22, no. 3 (2018): 439-445. https://doi.org/10.1007/s12603-017-0947-7

[9] Shou, Zhao-yu, Meng-ya Li, and Si-min Li. "Outlier detection based on multi-dimensional clustering and local density." *Journal of Central South University* 24, no. 6 (2017): 1299-1306. https://doi.org/10.1007/s11771-017-3535-4

[10] Abbas, Asad, Isma Farah Siddiqui, Scott Uk–Jin Lee, Ali Kashif Bashir, Waleed Ejaz, and Nawab Muhammad Faseeh Qureshi. "Multi-objective optimum solutions for IoT-based feature models of software product line." *IEEE Access* 6 (2018): 12228-12239. https://doi.org/10.1109/ACCESS.2018.2806944

[11] Kirimtat, Ayca, and Ondrej Krejcar. "Multi-objective optimization at the conceptual design phase of an office room through evolutionary computation." In *Recent Trends and Future Technology in Applied Intelligence: 31st International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2018, Montreal, QC, Canada, June 25-28, 2018, Proceedings 31*, pp. 679-684. Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-92058-0_65

[12] Kirimtat, Ayca, Ondrej Krejcar, Berk Ekici, and M. Fatih Tasgetiren. "Multi-objective energy and daylight optimization of amorphous shading devices in buildings." *Solar Energy* 185 (2019): 100-111. https://doi.org/10.1016/j.solener.2019.04.048

[13] Stewart, Geoffrey, and Mahmood Al-Khassaweneh. "An implementation of the HDBSCAN* clustering algorithm." *Applied Sciences* 12, no. 5 (2022): 2405. https://doi.org/10.3390/app12052405

[14] McConville, Ryan, Raul Santos-Rodriguez, Robert J. Piechocki, and Ian Craddock. "N2d:(not too) deep clustering via clustering the local manifold of an autoencoded embedding." In *2020 25th international conference on pattern recognition (ICPR)*, pp. 5145-5152. IEEE, 2021. https://doi.org/10.1109/ICPR48806.2021.9413131

[15] Baraković, Sabina, Jasmina Baraković Husić, Dardan Maraj, Arianit Maraj, Ondrej Krejcar, Petra Maresova, and Francisco Jose Melero. "Quality of life, quality of experience, and security perception in web of things: An overview of research opportunities." *Electronics* 9, no. 4 (2020): 700. https://doi.org/10.3390/electronics9040700

[16] Baraković, Sabin a, Jasmina Baraković Husić, Joost Van Hoof, Ondrej Krejcar, Petra Maresova, Zahid Akhtar, and Francisco Jose Melero. "Quality of life framework for personalised ageing: a systematic review of ICT solutions." *International journal of environmental research and public health* 17, no. 8 (2020): 2940. https://doi.org/10.3390/ijerph17082940

[17] Greffin, Klara, Holger Muehlan, Neeltje van den Berg, Wolfgang Hoffmann, Oliver Ritter, Michael Oeff, Georg Schomerus, and Silke Schmidt. "Setting-sensitive conceptualization and assessment of quality of life in telemedical care—study protocol of the Tele-QoL project." *International Journal of Environmental Research and Public Health* 18, no. 19 (2021): 10454. https://doi.org/10.3390/ijerph181910454

[18] Kasar, Kadriye Sayin, and Emine Karaman. "Life in lockdown: Social isolation, loneliness and quality of life in the elderly during the COVID-19 pandemic: A scoping review." *Geriatric Nursing* 42, no. 5 (2021): 1222-1229. https://doi.org/10.1016/j.gerinurse.2021.03.010

[19] Lwin, Ko Ko, and Yuji Murayama. "Modelling of urban green space walkability: Eco-friendly walk score calculator." *Computers, Environment and urban systems* 35, no. 5 (2011): 408-420. https://doi.org/10.1016/j.compenvurbsys.2011.05.002

[20] Komárková, Lenka, Petra Marešová, Peter Pažitný, Daniela Kandilaki, Filip Zemek, and Zuzana Truhlářová. "Healthcare Worker Capacities in Social Services for Elderly Care: Qualification and Demographic Structure." (2021). https://doi.org/10.36689/uhk/hed/2021-01-032