2024 European Conference on Computing in Construction
Chania, Crete, Greece
July 15-17, 2024

IDENTIFING REFERENCE DISTRICT BYMEANS OF MACHINE LEARNING AND OPEN-SOURCE DATA

Thi Thu Ha Dam[1], Maxim Shamovich[1], Julia Schlueter[2], Jannes Kruse[2]
Elisabeth Beusker[2], Jérôme Frisch[1] and Christoph van Treeck[1]

[1]Institute of Energy Efficiency and Sustainable Building (E3D), RWTH Aachen University, Germany
[2]Teaching and Research Field for Real Estate Development (IPE), RWTH Aachen University, Germany

Abstract

This study addresses neighborhood-level sustainability in urban development as a critical lever in Germany's energy transition goals. It explores identifying reference districts through Geographic Information System (GIS)-based machine learning (k-means algorithm) and public data, utilizing clustering methods to analyze spatial and socio-infrastructural metrics. The methodology yields significant insights into district definition and characterization, integrating technical and human understanding of urban dynamics. The findings highlight the importance of attribute selection in neighborhood classification and extend beyond mathematical validation to include social context comprehension. The developed technique is applied to a case study involving the city of Aachen.

Introduction

In the energy transition, particularly within the building sector, renewable energy systems are crucial for a climate-neutral energy supply, with heating and hot water consumption constituting a significant portion of Germany's building energy use (BMWK-Bundesministerium für Wirtschaft und Klimaschutz, 2022). The traditional, fossil-based energy infrastructure often limits the implementation of innovative, demand-driven thermal solutions (Acksel et al., 2017). District solutions are pivotal in optimizing energy potential and achieving economies of scale through sector coupling (Reicher et al., 2020). This approach integrates energy-saving strategies, efficiency, and renewable use, considering building characteristics, user behavior, and technical infrastructures (Többermann and Yu, 2021).

However, municipalities' current practices in urban-scale energy system planning are fragmented, often leading to slow and inefficient implementation (Boenigk et al., 2019). To address this, the paper advocates for establishing 'reference districts' as a means to streamline and standardize energy solutions, facilitating their application within and across various municipalities. Therefore, the paper discusses the challenges in defining district boundaries, especially in existing, evolving neighborhoods, and proposes using GIS and machine learning techniques on public datasets to identify and classify districts based on diverse attributes. This approach aims to enhance the scalability and effectiveness of energy solutions within existing infrastructures.

The concept of a neighborhood or district, lacking a universal definition, is interpreted variably across urban planning, cultural, and social perspectives (Feldmann, 2009). In urban planning, districts are defined by physical attributes such as building density, types, ages, and locations, including aspects like energy-efficient urban aggregation (Reicher, 2013; Reicher et al., 2020). Socially, they are important spaces for daily life, social interaction, and identity formation, embodying a range of physical to historical dimensions, thus influencing individual and communal life (Schnur, 2008). Neighborhood sizes vary, ranging from large housing estates to small settlements, with manageability and social identification as key criteria, typically not exceeding 20,000 residents (Bundesinstitut für Bau-, Stadt- und Raumforschung, 2012; Mehnert and Kremer-Preiß, 2014). Neighborhood attributes encompass a diverse spectrum including functional, spatial, socio-demographic, technical, economic, cultural-historical, and legal aspects, with differentiating factors such as natural boundaries, architectural typology, social structure, and community spaces (Malottki et al., 2013).

In case of newly planned districts, specific demarcation and energy zoning are simple, as they are planned in advance and meet current energy standards. In the case of existing districts, however, the question of boundaries arises because they are continuously developed, changed, and have grown over time. The question is, how can boundaries be defined in order to utilize energy potential in the context of the district?

For the practical application of district energy solutions, it is vital to recognize the diversity among district types. This paper puts forth a strategy for establishing neighborhood boundaries by employing georeferenced information combined with machine learning techniques on accessible datasets. Such techniques enable the amplification of energy solutions in pre-existing infrastructure, by creating reference districts defined by specific attributes. These can range from singular features like the type of buildings to an amalgamation of several traits. Consequently, this aids in the convenient replication of district energy solutions in comparable neighborhoods, resulting in a wide spectrum of reference districts, each marked by its distinct set of characteristics.

Related Research

This literature review explores statistical approaches, machine learning, and GIS in urban planning, emphasizing their role in energy evaluation and spatial analysis. It covers various approaches to mapping urban energy properties, classifying urban areas, and developing tools for sustainable urban development, highlighting the advancements and applications of these technologies in urban morphology analysis.

In the field of urban energy evaluation, several key projects and methodologies have emerged. The UrbanReNet project by Dettmar et al. (2020) systematically mapped the energetic properties of urban areas, analyzing prototypical elements of urban architecture and spaces for energy and structural values. This approach, combining urban morphology and land use, allows for a representation of Germany's building stock. It integrates qualitative and quantitative methods for evaluating energy generation, storage, and networking, leading to the development of mathematical models and software tools for neighborhood-level energy supply concepts.

GIS-based analyses have been pivotal in understanding energy dynamics in urban areas. März (2016) employed GIS-based multi-criteria decision analysis to locate neighborhoods at risk of energy poverty, focusing on space heating. Alpagut et al. (2021) and Quénéhervé et al. (2018) used GIS to optimize land use for solar production and its local energy impact.

In identifying and classifying settlement areas, Jochem et al. (2018) utilized geo-based vector data and machine learning to classify residential settlements in Afghanistan, while Gonzalez et al. (2020) applied deep learning to identify urban building typologies. Arribas-Bel et al. (2021) and Perez et al. (2020) explored urban cluster analyses using modified DBSCAN and Bayesian Networks, respectively, to analyze building types and urban functions.

Regarding city planning and administration, considering neighborhood dynamics, Photis (2012) developed the SPiRAL algorithm for redistricting electoral districts. López-Moreno et al. (2022) introduced a GIS-based approach for classifying residential areas in Madrid, aiding in energy-efficient urban renewal strategies. Similarly, Kelm et al. (2019) applied a semi-automatic approach using official geobase data in North Rhine-Westphalia (NRW) to create block structures for city planning and administration, demonstrating the growing importance of technology and data in urban planning and energy management.

Our methodology's choice of attributes for analysis was guided by this literature review on urban neighborhood characteristics, which highlighted key socio-infrastructural and energy-related factors to include in our study.

Methodology

Data Sources

In the process of identifying various reference districts in Aachen, NRW, a robust data foundation is essential. To conduct a comprehensive analysis, various public data sources were utilized in this study, with the main focus on the city of Aachen, located in western Germany. An overview of the freely available data sources used can be found in Table 1.

A central dataset titled ALKIS Real Estate Cadastre provides detailed information about plots and buildings in NRW. According to the Surveying and Cadastre Law of the state of NRW (VermKatG NRW), it offers extensive information about the structure and identification of plots and buildings. The cadastre, in which these data are recorded, offers a wealth of information, including geometric and geographical data, usage, size, and development of each plot. Additional datasets were sourced from the OpenGeodata.NRW portal, which includes, among others, information on general and psychiatric hospitals, was last updated in June 2023. Complementarily, locations of childcare centers (KiTas) and schools were also incorporated into the analysis, as they represent crucial social spaces within a neighborhood.

Energy-related data, such as the location of renewable energy sources or potential areas for renewable energy installations, is sourced from the Energieatlas NRW. It is collected and provided by the State Agency for Nature, Environment, and Consumer Protection of NRW.

The Approach

In this section of our study, we describe the methodology adopted for analyzing reference districts through a bi-fold approach focusing on socio-infrastructural factors and energy sources. The objective was to leverage the clustering of various public datasets to gain insights into neighborhood structures.

Multiple variables are considered to provide a comprehensive view of the various characteristics and dynamics within a neighborhood. This is achieved by first creating a so-called feature matrix. A feature matrix is a table where each row corresponds to an object (in this case, a building), and each column represents a feature or characteristic of that object (e.g., the distance to churches). In this context, the features represent urban planning factors. Each entry in this matrix denotes the value of the corresponding building feature. Once the feature matrix is established, a clustering algorithm can be applied. The first clustering process aims to identify districts that show similarities in terms of the features captured in the matrix. After the initial clustering, in which districts are identified, the dataset is enriched with additional information regarding the building type and specific and absolute heating demands from the Wärmekaster NRW dataset. Each district is characterized by the total amount of contained buildings, the distribution of building types, and the district's aggregated heating demand. Districts with similar feature values are grouped into the same cluster, resulting in a reference district, while those with significantly different values are classified into different clusters. This approach enables a multidimensional examination of the neighborhoods, allowing us to uncover hidden relationships and patterns between neighborhoods that might be overlooked in isolated analyses. Two different feature sets are considered and are described in the following sections.

Table 1: Overview of the utilized freely available data sources

Dataset	Description	Source
ALKIS Real Estate Cadastre	Buildings, Plots	Bezirksregierung Köln (2024)
Geoportal.NRW	Locations of Hospitals and Schools	Lanuv NRW (2024b)
Wärmekataster	Building Category, Building Type, Heat Demand	Lanuv NRW (2024c)
Open Street Map	Streets, Religious Institutions, Parks, Supermarkets, Car-sharing, Restaurants	OpenStreetMap Contributors (2024)
Solarkataster	Locations of Rooftop and Open-space PV	Lanuv NRW (2024a)
Energieatlas NRW	Biomass, Wind power plants, Hydropower plants, Lignite, Natural gas, Sewage gas, Mineral oil, Hard coal, Mine gas, Landfill gas	Bezirksregierung Köln (2024)

Socio-Infrastructural Approach

Various socio-infrastructural data were analyzed to examine a neighborhood's social structure. These will be described in the following.

In this study, building density is understood as the distance between individual buildings, which can provide insights into population density and the intensity of residential developments in a neighborhood. Typically, buildings in urban areas are constructed in close proximity to each other, whereas in rural areas, the distance between buildings tends to be greater and more variable. For instance, town or terraced houses often exhibit uniform distances from one another.

Considering the distance of buildings to religious institutions might reveal social and urban structural patterns. This concept dates back to ancient and medieval times, when churches, temples, and even city halls evolved as focal points of urban planning (Kaupp (2022)).

Furthermore, green spaces enhance the quality of life for adjacent residents, offering areas for both social interaction and individual recreation. It is anticipated that the proximity to parks influences the likelihood of residents utilizing these areas for leisure activities, potentially shaping local social and cultural boundaries.

In the state of NRW, according to the statewide hospital plan, accessibility to hospitals within a 20-minute drive should be guaranteed for 90% of its citizens (Ministerium für Arbeit Gesundheit und Soziales des Landes Nordrhein-Westfalen, 2023). In this context, the question arises whether the location of buildings and, thus, their proximity to hospitals plays a role in defining neighborhoods. Consequently, distance was selected as a relevant attribute for investigating this relationship. Additionally, the locations of schools and childcare centers might play a role in the social delineation of neighborhoods. Children often attend the school or childcare center closest to their residence. Areas with such facilities are typically characterized by family-friendly infrastructure, including playgrounds and similar amenities. Hence, the distance to playgrounds is also considered a parameter for investigation.

Moreover, local businesses could influence the character of neighborhoods. The distance to restaurants, for instance, might indicate the liveliness of an area. Accordingly, distances to restaurants and supermarkets were included in the study.

The spatial proximity to office buildings can indicate places of work and potentially offer new urban planning insights. Office buildings, often part of non-residential structures, tend to concentrate in specific areas along with other non-residential structures.

Furthermore, the structure of districts is often shaped by the traffic route network. Examining this factor could reveal how street layouts influence the formation of neighborhoods. In urban areas, car-sharing services are increasingly spreading. Analyzing the locations of such car-sharing facilities could provide indications of neighborhood boundaries.

In calculating distance as a feature, the shortest distance to the destination, for example, the nearest school, was determined. For streets with a speed limit of 50 km/h, the shortest orthogonal distance from a building to the street was chosen as the feature.

Table 2: Clustering attributes for Socio-infrastructural approach

Building density	Car-sharing	Religious institutions
Office buildings	Residential buildings	Childcare centers
Playgrounds	Restaurants	Supermarkets
Schools	Parks	Hospitals
Roads (max. 50 km/h)		

The complete set of attributes relevant to the Socio-Infrastructure Model is detailed in Table 2.

Energy Approach

In addressing the energy aspect of our study, we focus on the significant shift from fossil fuels to renewable energy sources, a strategic move by the Federal Government to counteract climate change. This transition is particularly pertinent in NRW, Germany's most populous state, which boasts a wide array of energy sources. These range from traditional fuels like lignite and mined gas to innovative renewable technologies, including biomass and photovoltaic (PV) energy generation. The pivot towards renewable sources signifies the phasing out of fossil fuels, underscoring the need to ensure continuous access to energy sources for the future. This is vital for maintaining an efficient energy supply chain. The crux of our energy approach is the development of an urban model designed to map out the distribution of various neighborhoods based on energy sources within urban settings. A key aspect of this model involves analyzing the proximity of buildings to diverse energy carriers, thereby identifying potential areas for the integration of renewable energy solutions based on the density of energy sources. By mapping the city's proximity to these energy sources, our model not only highlights Aachen's current energy infrastructure but also provides a blueprint for enhancing renewable energy uptake. This approach aligns with national efforts to transition towards a more sustainable and environmentally friendly energy mix, reflecting a commitment to reducing carbon emissions and combating climate change. The following attributes were considered for the clustering:

- Roof-surfaces PV
- Open-space PV
- Wind power plants
- Hydroelectric power plants
- Sewage gas
- Biomass availability

Integrating these two approaches enables us to construct a multidimensional description of urban neighborhoods, highlighting both challenges and opportunities in socio-infrastructural development, energy efficiency, and renewable energy utilization. Subsequently, districts from both approaches are enriched with data on building types and heating demand, resulting in a comprehensive dataset encompassing all identified districts and their attributes. An extra clustering step is then conducted to define reference districts, enabling the assignment of district types from the social infrastructure and energy perspectives.

After careful consideration of various algorithms, the k-means clustering method was chosen based on its efficiency in processing large datasets and its simplicity in implementation. Furthermore, the main hyperparameter k, the number of clusters, is highly interpretable, allowing expert knowledge to inform the algorithm tuning process. While DBSCAN was considered for its density-based clustering capabilities, it is earmarked for future work; however, it was discarded for the first proof of concept due to its complex parameter tuning.

Further, to evaluate the effectiveness of the k-means algorithm in producing meaningful clusters, we employ the Silhouette method. The Silhouette method is a popular evaluation technique, offering a quantifiable measure of cluster cohesion and separation, providing a robust justification for the selected number of clusters compared to other methods like the Elbow Method. It provides a succinct graphical representation of how well each object has been classified. The Silhouette score ranges from -1 to +1, where a high value indicates that the object is well-matched to its own cluster and poorly matched to neighboring clusters. This method is particularly useful in determining the optimal number of clusters and in assessing the quality of the clusters formed by the algorithm. A hyperparameter study was conducted using the Silhouette method for k-values ranging from 6 to 70. The minimum number of clusters was determined based on the number of the seven main administrative districts in the city of Aachen. The choice of the upper limit of 70 was also informed by the results of the Social Development Plan of the city of Aachen (Stadt Aachen, 2020).

Through this methodological approach, we aim to achieve a robust clustering of the given data to find definable reference districts, providing valuable insights into the underlying patterns and relationships within the dataset.

The study used the High-Performance Computing Cluster (HPC) at RWTH Aachen University for intensive computations. Due to QGIS's limitations in processing large datasets and tuning hyperparameters, it was replaced by Python for analysis, while QGIS was used only for visualization. The research employed Python libraries like scikit-learn for machine learning, GeoPandas and Pandas for data management and feature selection, and Fiona for loading geo-referenced data.

Results

Socio-infrastructural Approach

The methodology focussing on the socio-infrastructural attributes enabled us to delineate 54 unique districts, each representing a specific typology. The resulting cluster map can be seen in Figure 1. It can be observed that the larger clusters tend to be near and around the city center, which can be explained by the higher density of social infrastructure in the historical urban core.

Our analysis revealed significant differentiation among the clusters, particularly in terms of their spatial distribution and proximity to key urban infrastructures. For instance, significant variations were observed when examining the distribution of average distances of all clusters to residential buildings. Some clusters were characterized by notably shorter distances to residential areas, suggesting a predominance of residential use within those districts. This observation hints at the potential functional specialization of neighborhoods, where some are more residentially-focused while others may serve different urban functions.

Further examination of the average distance to religious in-

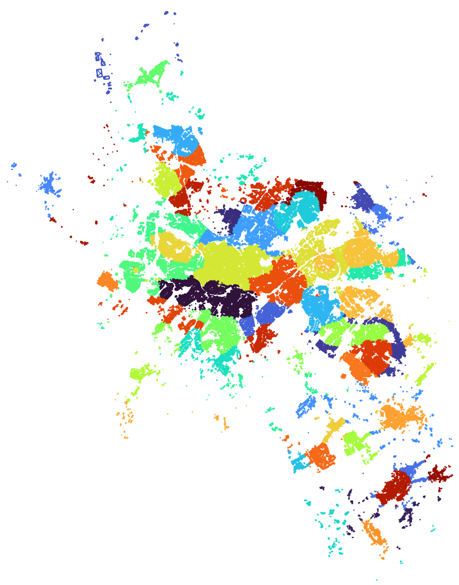

Figure 1: Clustering result of the Socio-Infrastructural-Model: 54 districts were identified

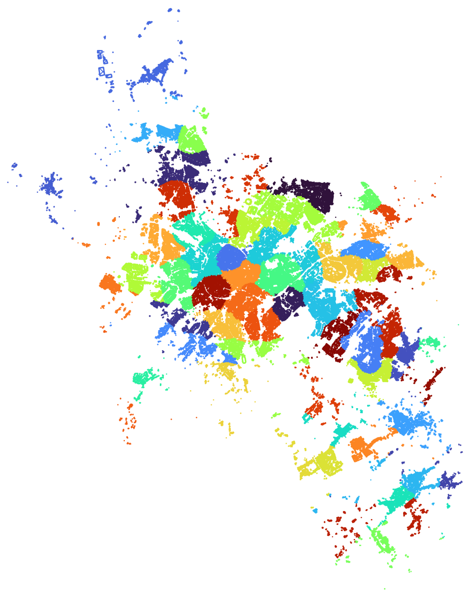

Figure 2: Clustering result of the Energy-Model: 58 districts were identified

stitutions across clusters also highlighted distinct patterns. Certain clusters demonstrated closer proximity to these institutions, reflecting their potential role in the social and cultural fabric of those neighborhoods. This differentiation in proximity to key urban elements underscores the diverse character of urban districts and their unique contributions to the city's overall structure.

Additionally, the analysis of the distribution of average distances for various attributes revealed nuanced differences between clusters. While some attributes showed a relatively linear distribution across clusters (e.g., distance to healthcare facilities), others exhibited significant variability (e.g., distances to childcare centers).

The observed variation indicates that specific urban characteristics might play a more significant role in distinguishing neighborhood types. There's a notable disparity in the number of buildings per cluster; for instance, clusters in the city center encompass over 6000 buildings, whereas those on the city's outskirts may have as few as 65 buildings. Additionally, there's a trend where the proportion of non-residential buildings in a cluster tends to rise with its distance from the city center. Meanwhile, the heating demand seems to be proportional to the cluster size.

Energy Approach

While looking at the different locations of energy resources in the city of Aachen, a total of 58 different clusters were found, with a Silhouette value of 0.38. Figure 2 shows the different clusters based on the location of various energy sources.

The clusters, each representing a unique combination of energy source proximity, were visually represented to provide insights into the spatial distribution of energy infrastructure across the urban fabric. This visualization underscores the variance in access to different energy sources across neighborhoods, such as biomass producers and PV installations. This examination revealed notable differences in how closely neighborhoods are situated to renewable energy sources, highlighting areas with potential for further development of sustainable energy solutions.

In particular, the study found varying degrees of proximity to biomass energy producers and rooftop PV systems across the clusters. Some clusters demonstrated close proximity to biomass sources, suggesting an emphasis on bioenergy utilization, while others were characterized by greater distances, indicating potential areas for increased biomass energy integration. Similarly, the examination of distances to PV installations provided a snapshot of the penetration of solar energy within urban areas, identifying clusters that could benefit from enhanced solar energy deployment.

In contrast to the first method, the range of building distribution per identified cluster is significantly narrower, with the highest count being 2334 buildings and the lowest at 70. Additionally, the total heating demand correlates less with the number of buildings and more with the types of buildings present. The distribution of building types per cluster also exhibits a less steep gradient compared to that observed in the first approach.

This clustering analysis sheds light on the intricate relationship between urban form and energy infrastructure, offering a foundation for targeted urban planning and energy policy development. By pinpointing districts with specific energy characteristics, policymakers and planners can tailor strategies to optimize energy efficiency and expand renewable energy use within urban ecosystems. This approach not only contributes to the sustainability of urban areas but also supports broader goals of energy transition and climate change mitigation.

Reference Districts

Integrating districts from socio-infrastructure and energy methods into a unified clustering process resulted in the identification of six unique reference district types. These types were defined by key characteristics: the total number of buildings in each cluster, the variety of building types, and the total aggregated heating demand. This comprehensive overlay of districts from both methods facilitated their categorization into one of the six reference types, as depicted in Figure 3. The clustering effectiveness, measured by the Silhouette score, was 0.433 for k=6.

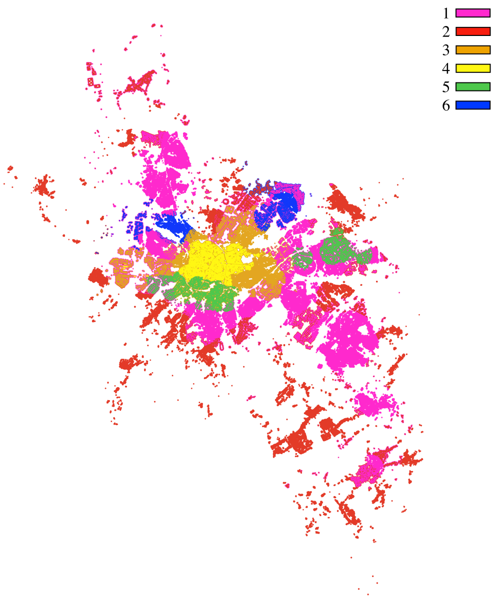

Figure 3: Results of the Reference Districts Clustering: Overlay of Socio-Infrastructural and Energy Model Districts

On the map, it is clearly visible that the reference district types form an orbicular pattern around the city center. Table 3 summarizes the mean distributions of each building type and specific heating demand for the six reference district types.

Cluster number four represents Aachen's urban center and is the city's most densely populated region, which is evident from the high number of buildings in this reference district. Characterized predominantly by large multi-family housing (BMFH), this cluster has the highest specific heating demand among all clusters, significantly exceeding others. In contrast, cluster two, encompassing the city's outskirts, has the lowest heating demand. Clusters one and six share striking similarities, not only in residential building distribution but also in heating demand and location, as depicted in Figure 3. However, there are significant differences in their absolute heating demands. The total demand of cluster six is nearly 1.5 times higher than that of cluster one.

A detailed examination of Table 3 highlights the primary emphasis of the clustering process. It shows that each identified reference district type is predominantly characterized by a single building type.

Discussion

The clustering approach uses a feature matrix aimed at a holistic approach, considering multiple attributes simultaneously. In this study, while defining two base models for identifying typological neighborhoods using public datasets, it was crucial to select relevant attributes, as irrelevant ones could distort outcomes. Selecting specific attributes for clustering still remained a subjective challenge. The approach incorporated biases related to socio-infrastructural aspects to provide insights into patterns that may be prevalent across other communities within NRW and thereby acknowledging the influence of preconceived constructs of the attributes. The results demonstrated that not all attributes cluster effectively, with some showing little variance across neighborhoods, highlighting the need for careful attribute selection in urban area analysis due to the dynamic nature of urban development and infrastructure changes. Additionally, increasing dimensions complicated clustering without compromising quality, often resulting in indistinct clusters.

The chosen k-means algorithm necessitates predefining the number of clusters (k), with determining the ideal number being a significant challenge. Computation efficiency requires setting a specific range for k, as demonstrated in the Aachen case study. However, estimating this range becomes increasingly complex for larger or more diverse geographical areas, such as the whole state of NRW, which encompasses multiple cities and rural areas, the latter being outside the scope of this study. The utilization of the

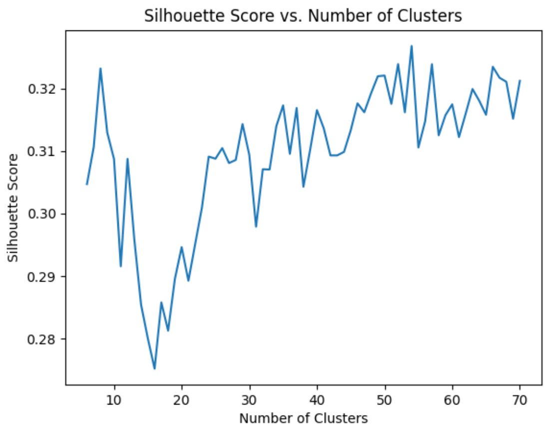

Figure 4: Silhouette score for k-means clustering of Socio-Infrastructural-Model for different k

Silhouette score to identify the optimal 'k' for k-means clustering was a critical component of our analysis. This evaluation method, grounded in mathematical principles, provides a quantifiable measure of cluster coherence and separation. However, it is essential to note the distinction between the mathematical assessment of cluster quality and the perceived social characteristics of neighborhoods. While the mathematical evaluation offers a systematic ap-

Table 3: Mean distribution of cluster attributes: Single-Family-Dwelling (SFD), Big-Multi-Family-House(BMFH), Terraced House (TH), Multi-Family-House (MFH), Non-Residental Buildings (Non.-Res), specific heating demand (demand)

Cluster	SFD	BMFH	HR	MFH	TH	Non-Res.	demand $\mathrm{kW\,h\,m^{-1}\,a^{-1}}$
1 pink	490.2	106.17	0.5	80.66	314.37	169.79	156305.82
2 red	190.22	33.87	0.1	34.66	67.00	93.06	56945.96
3 orange	232.90	728.60	1.30	205.30	257.90	535.00	269755.08
4 yellow	411.0	2451.00	4.0	825.0	545.00	2370.00	956452.06
5 green	787.50	746.00	1.0	351.00	915.00	587.50	472010.96
6 blue	355.75	164.25	5.25	101.75	178.25	353.25	157519.04

proach to determining cluster quality, it may not capture all aspects relevant to interpreting these clusters effectively. Moreover, it was observed that the graph displayed comparably high Silhouette scores with significantly fewer clusters (Figure 4). This phenomenon suggests that a more compact cluster formation can provide a similarly coherent and well-separated structure as configurations with a larger number of clusters. This finding is particularly illuminating, as it implies that optimal clustering does not necessarily equate to maximizing the number of clusters but rather to identifying a configuration that balances cluster coherence with the complexity of the model. This discrepancy underscores the importance of not solely relying on quantitative metrics for interpreting cluster results. An understanding of the urban structure, historical developments, and other pertinent factors should also be incorporated. Such a holistic approach ensures that the interpretation's significance is not merely confined to numerical values but is enriched by a comprehensive context of neighborhood structures.

Conclusion and Future Work

This study yielded significant insights into the feasibility of neighborhood classification studies while highlighting opportunities for future research. It showcased an innovative approach to urban planning and energy management by integrating socio-infrastructural and energy data to identify reference districts within Aachen, NRW. The cluster analysis conducted for Aachen revealed specific structural and characteristic patterns within its neighborhoods. By employing clustering techniques on public datasets, the research illuminated the complex interplay between urban form, energy infrastructure, and social dynamics, contributing valuable insights into sustainable urban development and the energy transition. The methodology demonstrated the potential of using machine learning and GIS technologies to dissect and understand the multifaceted characteristics of urban areas. However, the study also highlighted several challenges, including the subjectivity in selecting clustering attributes, the difficulty in managing and integrating large and diverse datasets, and the computational demands of processing extensive urban data. Applying the same analysis to the entire state of NRW would likely yield different results due to its diverse composition of cities, rural areas, and villages, each with unique structures and attributes, requiring spatial understanding and various validation methods. Investigating various validation methods could also comprehensively assess cluster result quality. A systematic examination of the weighting of individual characteristics may enable a more nuanced, context-specific classification, acknowledging the varying impacts of different features on neighborhood formation. These challenges underscore the need for further refinement of the methodology and the exploration of more advanced machine learning techniques to enhance the precision and applicability of the findings. Additionally, integrating socio-economic or environmental indicators could better represent neighborhood characteristics and contribute to developing more comprehensive models. Including the performance and impacts of energy producers on the power and heating networks and building energy inputs in the study of energy neighborhoods could offer a more holistic view.

Acknowledgments

We gratefully acknowledge the financial support by BMWK, promotional reference 03EWR010B.

References

Acksel, D., Huenges, E., and Kastner, O. (2017). Wärmewende am Beispiel Quartier: Ein Beitrag zur Energiewende.

Alpagut, B., Lopez Romo, A., Hernández, P., Tabanoğlu, O., and Hermoso Martinez, N. (2021). A GIS-Based Multicriteria Assessment for Identification of Positive Energy Districts Boundary in Cities. Energies, 14(22):7517.

Arribas-Bel, D., Garcia-López, M.- ., and Viladecans-Marsal, E. (2021). Buildings and cities: Delineating urban areas with a machine learning algorithm. Journal of Urban Economics, 125:103217.

Bezirksregierung Köln (2024). ALKIS NW Grundrissdaten. Accessed: 2023-05-01.

BMWK-Bundesministerium für Wirtschaft und Klimaschutz (2022). Unsere Energiewende: sicher, sauber, bezahlbar. Accessed: 2024-01-29.

Boenigk, N., Cantos, E., Eyerich, L., Kradolfer, A., Müller, A., Müller, I., and Schwalbe, A. (2019). Komm:mag: Das Jahresmagazin zu erneuerbaren Energien in Kommunen.

Bundesinstitut für Bau-, Stadt- und Raumforschung (2012). Bestand und städtebauliche Bedeutung - neue Stadtquartiere.

Dettmar, J., Drebes, C., and Sieber, S., editors (2020). Energetische Stadtraumtypen. Fraunhofer IRB Verlag.

Feldmann, P. (2009). Die strategische Entwicklung neuer Stadtquartiere: unter besonderer Berücksichtigung innenstadtnaher oder innerstädtischer, brachgefallener Industrieareale, volume 53. Immobilien-Manager-Verl.

Gonzalez, D., Rueda-Plata, D., Acevedo, A. B., Duque, J. C., Ramos-Pollan, R., Betancourt, A., and Garcia, S. (2020). Automatic detection of building typology using deep learning methods on street level images. Building and Environment, 177:106805.

Jochem, W. C., Bird, T. J., and Tatem, A. J. (2018). Identifying residential neighbourhood types from settlement points in a machine learning approach. Computers, Environment and Urban Systems, 69:104–113.

Kaupp, A. (2022). Corporate urban responsibility–Kirche in der Stadtentwicklung. In CSR und Kirche: Die unternehmerische Verantwortung der Kirchen für die ökologisch-soziale Zukunftsgestaltung, pages 261–272. Springer.

Kelm, T., Schonlau, M., Pitz, N., and Klein, U. (2019). Semiautomatisches Verfahren zur Ableitung von Baubloecken. Selbstverlag des Vereins CORP - Competence Center of Urban and Regional Planning, Wien, Österreich. Meeting Name: REAL CORP.

Lanuv NRW (2024a). Dachflächen-Solarthermie: Potentialdaten Solarkataster NRW. Accessed: 2023-08-01.

Lanuv NRW (2024b). Krankenhäuser. Accessed: 2023-06-01.

Lanuv NRW (2024c). Raumwärmebedarfsmodell NRW. Accessed: 2023-07-01.

López-Moreno, H., Núñez-Peiró, M., Sánchez-Guevara, C., and Neila, J. (2022). On the identification of Homogeneous Urban Zones for the residential buildings' energy evaluation. Building and Environment, 207:108451.

Malottki, C., Koch, T., and Vaché, M. (2013). Anforderungen an energieeffiziente und klimaneutrale quartiere (eq). Werkstatt: Praxis. Bundesministerium für Verkehr Bau und Stadtentwicklung. Bonn, Institut für Wohnen und Umwelt.

März, S. (2016). Identifikation kleinräumiger Hotspots der Energiearmut : Gis-gestützte Analysen zur Vulnerabilität von Quartiersbewohnern am Beispiel Oberhausen. In Schmitt, H. C., editor, Raummuster : Struktur, Dynamik, Planung, pages 101 – 119. Klartext-Verl., Essen.

Mehnert, T. and Kremer-Preiß, U. (2014). Ist-Analysen im Quartier. Handreichung im Rahmen des Förderbausteins 3.1. 1 „Projekte mit Ansatz zur Quartiersentwicklung "des Deutschen Hilfswerks.

Ministerium für Arbeit Gesundheit und Soziales des Landes Nordrhein-Westfalen (2023). Krankenhausplan Nordrhein-Westfalen 2022. die Strukturen müssen für die Menschen da sein, nicht die Menschen für die Strukturen! Accessed: 2023-08-21.

OpenStreetMap Contributors (2024). Open Street Map Daten Nordrhein-Westfalen. Accessed: 2023-07-01.

Perez, J., Fusco, G., Araldi, A., and Fuse, T. (2020). Identifying building typologies and their spatial patterns in the metropolitan areas of Marseille and Osaka. Asia-Pacific Journal of Regional Science, 4(1):193–217.

Photis, Y. N. (2012). Redefinition of the greek electoral districts through the applicaton of a region-building algorithm. European Journal of Geography, Volume 3(Issue 2):72–83.

Quénéhervé, G., Tischler, J., and Hochschild, V. (2018). Energiewende im Quartier–ein Ansatz im Reallabor. Bausteine der Energiewende, pages 385–405.

Reicher, C. (2013). Das (Stadt) Quartier, vom Umgang mit dem gebauten Raum und seinen dynamischen Parametern. Stadtquartiere. Sozialwissenschaftliche, ökonomische und städtebaulich-architektonische Perspektiven. Klartext, Essen, pages 197–209.

Reicher, C., Schmidt, A., and Hangebruch, N. (2020). Energieeffizienz und Quartier: Herausforderung Energieeffizienz im Quartier. Handbuch Energieeffizienz im Quartier: Clever versorgen, umbauen, aktivieren, pages 1–16.

Schnur, O. (2008). Quartiersforschung im überblick: Konzepte, Definitionen und aktuelle Perspektiven. Quartiersforschung: Zwischen Theorie und Praxis, pages 19–51.

Stadt Aachen (2020). Sozialentwicklungsplan Aachen. Accessed: 2023-08-24.

Többermann, I. J.-C. and Yu, I. Y. J. (2021). Energiewendepotentiale von Quartieren und Quartiersgrössen. Forschungsberichte, page 69.