The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume L-4/W2-2026
21st International 3D GeoInfo Conference 2026, 28–29 September 2026, Sofia, Bulgaria

Using AI to assign building archetypes to individual buildings for enhanced resolution and accuracy in urban digital twin applications

Matthias Betz[1], Robert Otto[1], Bastian Schröter[1]

[1] University of Applied Sciences Stuttgart, 70174 Stuttgart, Germany - (matthias.betz, robert.otto, bastian.schroeter)@hft-stuttgart.de

Keywords: Simulation, Machine Learning, Urban Digital Twin, Building Archetypes, CityGML

Abstract

Urban digital twins benefit from granular, building-level data, yet classifying city quarters into meaningful typologies remains a challenge. This paper presents a machine learning-based approach to automatically assign a building archetype to individual buildings using 3D CityGML data including building functions as the sole input. The archetypes range from detached single-family housing to industrial and business parks, and are defined by geometric properties and locational context rather than socio-economic factors, ensuring broad applicability across research domains. Because the method relies solely on widely available CityGML data, it is both portable and sector-agnostic, making it suitable for applications in energy planning, mobility research, and urban policy analysis.

The classification pipeline is built around the urban energy simulation platform SimStadt, which extracts per-building properties from CityGML files including building height, footprint, volume, storeys, roof type, usage, and statistically derived household and occupancy data. Crucially, the approach also incorporates neighborhood context by aggregating surrounding building characteristics within a 100-meter radius. These features feed a TensorFlow-based deep learning model trained on 111 manually labeled buildings. Applied to a test dataset of 17,039 buildings from the Stuttgart region, the model completed classification in just over two minutes. Validation results show strong overall performance, with a weighted average F1 score of 82%. These results are, however, a proof of concept rather than a conclusive assessment, given the limited size of the validation dataset. Next steps are an enhanced training and validation dataset or cross-validation with other approaches, among others.

1. Introduction

In recent years, the concept of urban digital twins has gained traction, as it allows the investigation of a wide range of city-related parameters from, for example, environmental, societal or economic areas (Zhu and Jin, 2025). The associated data points are most useful when geo-located and assessed on a local scale, as input and output parameter values vary even between city quarters or building typologies within one city, as has been shown for example for building heating options(Theile et al., 2022). A standardized classification of buildings into archetypes is thus important for a more granular data analysis as has been proposed by various authors (e.g. (Dettmar et al., 2020)).

To give a recent example, (Otto et al., 2026) describes a process to simulate future power distribution grid loads in urban areas by applying ramp-up projections for heat pumps, electric vehicles and rooftop PV on a single-building level. These ramp-ups are, however, dependent on the building archetype, with for example single-family homes having generally more space and decision power to install a wall box for EV charging or a heat pump, whereas denser multi-family buildings in inner-city areas often lack parking space for wall box installations and are more often supplied by district heating, suppressing heat pump penetration rates. A classification of an urban area into building archetypes and the provision of distinct technology ramp-up curves per such archetype thus allows for more accurate assessments of future power distribution grid loads than applying a standard ramp-up to every building. Linking technologies or other data to building archetypes can thus improve the climatic, transport or environmental assessments in urban digital twin applications (Li et al., 2020) (Zeynali et al., 2025) (Enteria et al., 2021).

In this work, we present a process that leverages machine learning (ML) routines to develop a model that assigns building archetypes to individual buildings within 3D building data in the CityGML file format for Germany. The presented work is to be understood as a proof of concept demonstrating the feasibility of the proposed process; a systematic (cross-)validation and further training runs with an extended labeled dataset are part of future work. As part of the input data, we account for the morphological characteristics of each individual building and its built environment within a 100 m radius. Both data points are derived automatically from the CityGML file, both during model training and during classification. Once the model is trained, a CityGML file containing the geometry and a building function denoting each building's usage are the sole area-specific input required for classification. All other parameters are derived from this base by using publicly available German census statistics for the household and occupancy estimates. The approach is thereby easily scalable to other world regions, if statistical information similar to Germany's is available and parametrized, as the CityGML format is widely used in urban digital twin applications. For the same reason, the presented work is technologically agnostic and can be applied to a range of research fields that benefit from the higher resolution provided by linking attributes relevant to the field to distinct building typologies.

2. Methodology

On a high level, the proposed approach requires three elements: first, a list of building archetypes that serves as a blueprint. Second, relevant and sector-agnostic information on the individual building level such as statistical data on occupancy rates that can be annotated to each building in the CityGML file used for training a model. Third, the choice of an appropriate ML model.

2.1 Definition of Building Archetypes

We define the following seven building archetypes in accordance with (Otto et al., 2026). In both works, building archetypes are defined by their geometric properties, whereas other parameters such as socio-economic factors are not considered. While this represents a simplification, it has the advantage that the AI-trained model requires only CityGML files and thus morphological data as input for classification:

- **BA1: (Detached) Single-family housing:** Single family and two family houses, as most often found in suburban and rural settings.
- **BA2: Multi-storey residential:** Houses with three to nine apartments in detached houses, describing a form of living predominantly found in urban or suburban settings.
- **BA3: City center residential:** Residential housing in or close to city centers. Mostly buildings with five or more apartments, three or more floor levels. Compared to BA2, BA3 buildings are typically attached in perimeter block developments, often feature commercial ground floor usages in their vicinity, and differ systematically in application-relevant properties as motivated in the introduction, e.g. the availability of district heating or the scarcity of parking space. They are thus treated as a separate archetype rather than a sub-category of BA2.
- **BA4: Large residential building:** Larger, detached building blocks or high-rise buildings with at least ten apartments per building. Typical examples include 1960s and 1970s social housing developments in the periphery of European cities, and more recently in Asia.
- **BA5: Mixed residential and commercial:** A typically inner-city type of buildings mixing ”light” industrial/craftsmen, commercial, and residential uses. This building type can have purely residential buildings in its surroundings. The defining criterion is a mix of usages within the same building, e.g., a commercial ground floor with residential storeys above, and a mix of primary building functions in each building's immediate neighborhood.
- **BA6: Urban commercial:** A type of building with predominantly commercial and ”light” industrial (exclusive) use that can be found close to city centers. Typical examples include office-only blocks or campuses. While of homogeneous use, the building type can have purely residential buildings in its neighborhood. In contrast to BA5, buildings of this archetype are used exclusively for commercial or industrial purposes, even if residential buildings are located nearby.
- **BA7: Industrial and business parks:** Industrial complexes as typically found in peri-urban settings, clearly separated from residential areas.

Terraced and row houses, as typically found in many European countries, are not treated as a separate archetype: predominantly single-family rows in suburban settings are assigned to BA1, while attached residential buildings in inner-city perimeter block developments fall under BA3. The row house (RH) type of the SimStadt building typology (see section 2.2) serves as an input parameter for the model, such that this information is available for classification.

As mentioned in the introduction, annotated data are on an individual building level, which means the assignment of archetypes is to be on the individual building level as well. However, for some building archetypes, notably to differentiate between BA5, BA6 and BA7, it is important to also consider the individual building's neighbors.

2.2 Data input

The general workflow from input data to ML model is shown in Figure 1. A core data base is the 3D building data in CityGML format, which is subsequently used by the urban energy simulation platform SimStadt[1]. The CityGML file contains information about the geo-referenced geometry of all buildings in a city or city quarter, plus annotated information that can be included as meta data in the CityGML file. This most often includes a building's usage function (see section 2.2.6 for more details) while other information such as the building's year of construction or the number of apartments per building are generally not included. For Germany, for example, all states provide CityGML data for the entire building stock as open data.

The CityGML file is then used as input by the urban energy simulation platform SimStadt to generate (meta) information on a single building level that is to be used as ML model training data. SimStadt has been steadily extended over the last years; its modular, workflow-driven design (Nouvel et al., 2015) facilitates diverse simulations and includes to date solar and photovoltaic potential assessments (Rodríguez et al., 2017), heat demand (Nouvel et al., 2014) and electricity demand (Köhler, 2019), emissions (Weiler et al., 2022), refurbishment scenarios, and urban water demand (Bao et al., 2020), among others. The modular design also enables SimStadt's workflows to often rely on common input files such as weather data and a database on building physics parameters, thus ensuring consistent results across multiple dimensions of urban energy systems and sustainability issues.

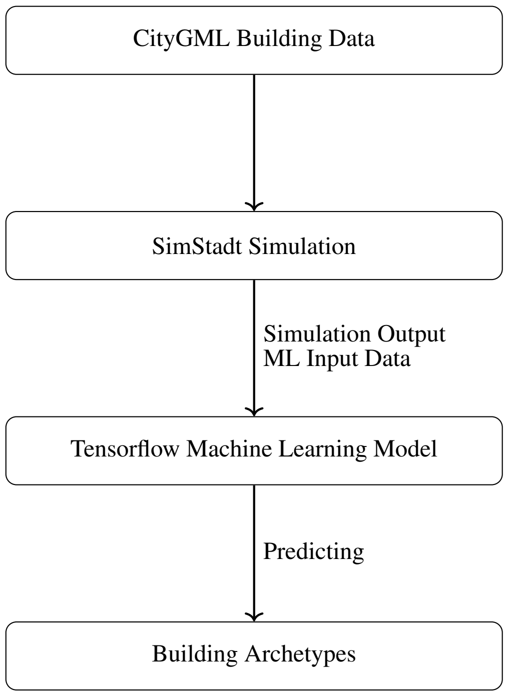

Figure 1. Data flow used for classifying building archetypes

SimStadt is here used for the whole classification process: it writes the required (meta) data into a temporary CSV file with

[1] Available at https://simstadt.hft-stuttgart.de/

each row representing one building. Next, SimStadt initiates a Python script that starts the machine learning process: it first reads the above-mentioned CSV file and inputs it into a *tensorflow* classification model (see also section 2.3) that produces a second CSV file as output containing the building archetype assignment per building in the classification as proposed in section 2.1.

For this process, the following parameters are assessed by SimStadt and annotated to the CityGML file as input/training data:

2.2.1 Building height The building height as measured in meters is calculated from the 3D geometry subtracting the lowest Z value from the highest Z value of any vertex, where the Z axis represents the height above sea level. For model training, and since it is preferable not to have too many building height values that differ only slightly, values are rounded to the nearest meter. For each meter, a distinct class was thus created.

2.2.2 Storeys While the number of storeys in a building correlates with its height, it can nevertheless provide additional information as it also takes roof shapes and potential basements into account. The storey calculation in SimStadt uses information on whether the roof shape allows for a heated attic or if there is information about the surrounding terrain to assume some of the geometry is a cellar. It also takes into account the average storey heights as they were typical in the respective year of construction. All of this information aside from the roof geometry - to determine if a roof is flat or not - is optional. When they are missing, general assumptions are made.

2.2.3 Building footprint The building footprint can be calculated from the CityGML data, which contains all relevant semantic information on the geometry and also declares which polygons are walls, roofs or ground surfaces. The footprint is thus determined by the area of all ground surface polygons. As with the building height, this value was rounded to the nearest square meter.

2.2.4 Building volume Similarly to the footprint, a building's volume is calculated from geometric information in the CityGML file. However, calculating volumes requires a "watertight" geometry, which is not always the case. In such cases, an object-aligned bounding box is used as an approximation for the volume and written to the temporary CSV file.

2.2.5 Building type As a default, one of five typologies is assigned to each building within SimStadt. These include SFH (Single Family House), MFH (Multi Family House), RH (Row House), AP (Large Apartment Buildings), HR (High Rise Buildings). This classification largely depends on the building height and on whether the building is detached or not: if a SFH building shares at least one wall with another building, it is considered a row house. Compared to the classification described in section 2.1, these five typologies are purely based on geometry and neglect information on a building's neighbors or usage type, among others. However, the information nevertheless serves as relevant and additional training data; each type is thus converted into a discrete boolean parameter as a separate input parameter. It is nevertheless important to clearly differentiate between these five legacy building types and the newly defined seven building *archetypes* described in section 2.1: while the former are geometry-based input parameters generated automatically within SimStadt, the latter represent the target classes of this work, additionally encoding usage and neighborhood information. The SimStadt-inherent legacy building type thus serves as one of several input parameters from which the building *archetype* is predicted.

2.2.6 Usage A key piece of information is a building's primary usage function, typically provided in each CityGML open data file as a code value per building. Codes can be either very detailed or rather general, we used the following level of detail for categorization:

1. Residential
2. Commercial
3. Industrial
4. Hospitality
5. Culture, Religion and Event
6. Leisure and Recreation
7. Other
8. Public
9. Agriculture and forestry
10. no building function

On this level of detail, the information is available in CityGML data sets as provided by most German states[2]. To reduce the number of categories for training, only the first four above-mentioned usage categories are used for classification. These are converted to discrete boolean input parameters in the temporary CSV file. That means the model is not able to classify buildings with a usage type other than the first four types.

2.2.7 Roof type CityGML's so-called *Level of Detail 2* data includes roof shapes, albeit without dormer or chimney details. Next to being relevant for, e.g., correctly dimensioning rooftop photovoltaic plants, roof shapes yield supplementary training data, since BA 5...7 (see section 2.1) typically do not feature tilted roofs. The information on whether a roof is flat or not is used as a boolean parameter.

2.2.8 Number of households or apartments per building and Occupants Information on the number of households or apartments and occupants in a building is not contained in the CityGML input data, but can be statistically derived from the heated area of a building. For this, a building's heated area is calculated in SimStadt and then split up into apartments using German census data (Landesämter, 2022), containing statistics for household sizes on a (100 x 100) meter grid scale, which are distributed over the heated area. This approach is only taken for residential buildings. This data is used to create an algorithm to distribute households and occupants for residential buildings. Since the heated area is derived from the 3D geometry including the number of storeys, the building height serves as implicit input here. This approach to distributing households and occupants to buildings is described in more detail in (Köhler et al., 2021).

2.2.9 Building surroundings For a more accurate classification, information on a building's neighbors is included in the training data, since the likelihood of a building being, e.g., of predominantly commercial usage increases with its neighbors being of that type, too. To accommodate for this, a circle with a 100 m radius is drawn around each building and geometrically intersected with the footprint area of the buildings within that circle. Depending on the usage types of these buildings, the footprint area is added to either a cumulative residential or commercial-industrial surrounding area. For this, the usage

[2] See for example here: `https://opengeodata.lgl-bw.de/#/(sidenav:product/lod2)`

categories *industrial*, *commercial*, *agriculture and forestry* and *hospitality* are merged into a *commercial-industrial* surrounding area for simplification. As this serves to determine what the surrounding area looks like, more usage types were included than we used as input data (see section 2.2.6). The resulting cumulative areas are normalized over all buildings of the training data: as an example, the building with the highest cumulative residential footprint area receives a value of 1.

2.3 Machine Learning

For model training, we use *Tensorflow* (Singh et al., 2020) as the machine learning framework, with a Keras Sequential (Gulli and Pal, 2017) deep learning algorithm applying two layers with 32 and 16 nodes and with a seven node output layer per building archetype. The 111 buildings of the learning dataset were labeled manually by one of the authors, based on aerial and street-level imagery, local knowledge of the Stuttgart area, and the attributes contained in the CityGML data. The buildings were selected such that all seven building archetypes are covered. Ambiguous cases were excluded from the dataset. A formal assessment of inter-annotator agreement, e.g. via Cohen's kappa, has not been performed. We acknowledge this as a limitation to be addressed when extending the labeled dataset. All decimal inputs are normalized, while the categoric inputs are converted to boolean inputs of 0 and 1 via a one-hot encoding.

The learning dataset is fitted with 150 epochs with an early exit if the training evaluation determines no improvements. It is split into 80% training and 20% validation data to evaluate learning performance. This results in a small validation dataset of 22 rows. For comparison, a random forest classifier was trained on the identical input data, reaching an accuracy of 63.6% on the same validation dataset, indicating that the neural network performs comparably.

The resulting deep learning model is applied to each row of the CSV dataset which represents one building. The resulting classification is then read back into SimStadt and inserted as attribute into the buildings where it can be used in further workflow steps.

Generally and due to geometric errors or missing input data such as an unknown function code, some buildings cannot be processed by SimStadt, so that no complete parameter set is available for them. For instance, a range of buildings are garages that have been deliberately excluded from the classification by not generating a row to classify as they should have the same archetype as the building they are associated with. To assign a building archetype to those buildings, the surrounding buildings are used to determine the most common archetype, which is then used for the original building as well. This is handled analogously to the principles mentioned in section 2.2.9; however, instead of summing up the footprint of buildings with a certain function, the previously assigned archetypes are used. In a test CityGML dataset from the region of Stuttgart with 17,039 buildings, 5,341 were not classified, of which 4,515 were garages and 323 canopies according to their building function code. This leaves just 503 or 3% possible classifications for which the required input data was not directly available or could not be generated from SimStadt, and which therefore could not be classified.

3. Results

It took the test dataset with 17,039 buildings 2 minutes and 4 seconds to be classified with a trained model on an Intel(R) Core(TM) i7-8850H @ 2.60GHz processor with the above-mentioned exceptions and non-classifiable buildings.

The confusion matrix in table 2 shows that all three BA5 buildings in the validation data were classified as BA6, resulting in precision, recall and F1 scores of 0.00 for this archetype; the model in its current state thus fails to identify the mixed-use archetype BA5 *Mixed residential and commercial*. All other building archetypes reached 100% precision. This results in an overall accuracy of 86% and a weighted average F1 score of 82% (see table 1). The small validation sample size results in a high uncertainty for both the F1 score (van Rijsbergen, 1979) and confusion matrix metrics (tables 1 and 2). Due to the limited dataset size, further validation steps such as cross-validation, repeated runs or confidence intervals have not been performed yet and are planned once the labeled dataset has been extended. The results presented here, in particular for building archetypes with a support of only two or three buildings, are thus to be interpreted with caution.

Class	Precision	Recall	F1-score	Support
BA1	1.00	1.00	1.00	5
BA2	1.00	1.00	1.00	3
BA3	1.00	1.00	1.00	3
BA4	1.00	1.00	1.00	3
BA5	0.00	0.00	0.00	3
BA6	0.40	1.00	0.57	2
BA7	1.00	1.00	1.00	3
Accuracy		0.86		22
Macro avg	0.77	0.86	0.80	22
Weighted avg	0.81	0.86	0.82	22

Table 1. F1 scores for the building archetype AI training

True \Pred	BA1	BA2	BA3	BA4	BA5	BA6	BA7
$BA1$	5	0	0	0	0	0	0
$BA2$	0	3	0	0	0	0	0
$BA3$	0	0	3	0	0	0	0
$BA4$	0	0	0	3	0	0	0
$BA5$	0	0	0	0	0	3	0
$BA6$	0	0	0	0	0	2	0
$BA7$	0	0	0	0	0	0	3

Table 2. Confusion matrix of the building archetype AI training

Figure 3 summarizes the correlations between the defined building archetypes and data inputs as described in section 2.2. As would be expected, it shows strong correlation between height and storeys, households/occupants and building archetypes: as one example, the building type SFH (type_EFH in the Figure) correlates strongly with BA1 (detached single-family housing), while buildings with flat roofs are more likely to be BA5 *mixed residential and commercial* or BA7 *industrial and business parks* than BA1-4. There are also correlations between a high number of storeys and households and archetype BA3 *city center*.

Figure 2 shows color-coded results for an exemplary city quarter in Stuttgart, Germany, with the bottom left indicating the respective building archetypes, also summarized in table 3. Similar assessments for at least other parts of Germany can be performed with the trained model for any city quarter that can be selected as a polygon of any shape from a larger CityGML file.

Figures 4 and 5 show on the left two city quarter maps from two different parts of Stuttgart to which the classification algorithm

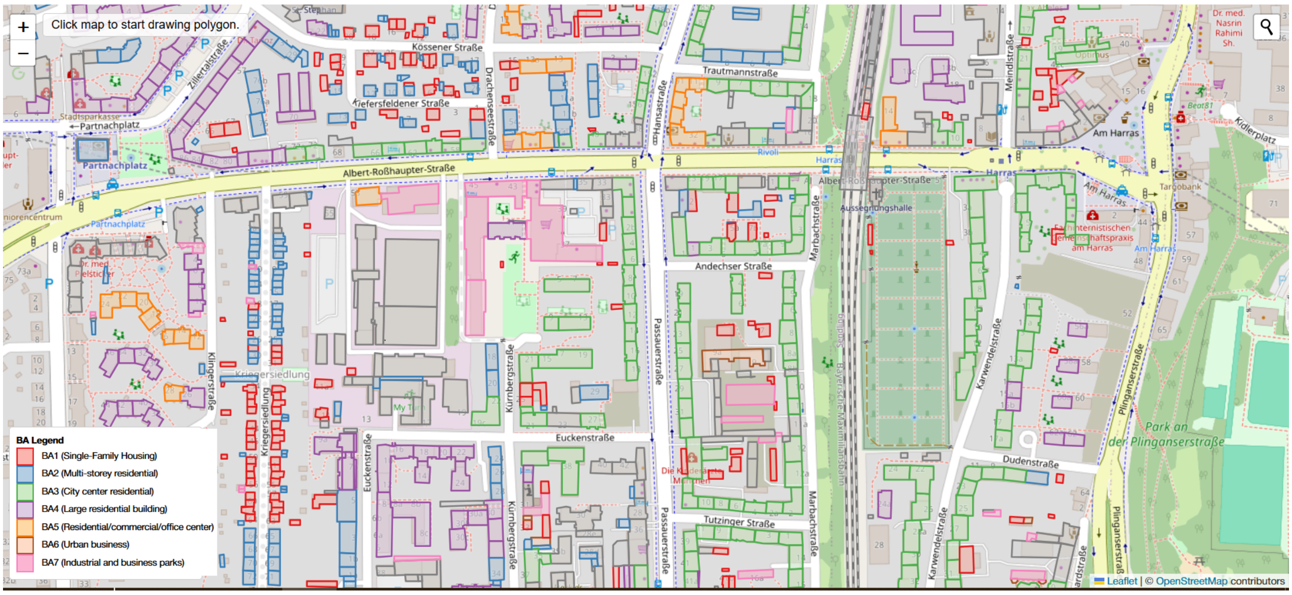

Figure 2. Building archetypes assigned to each building

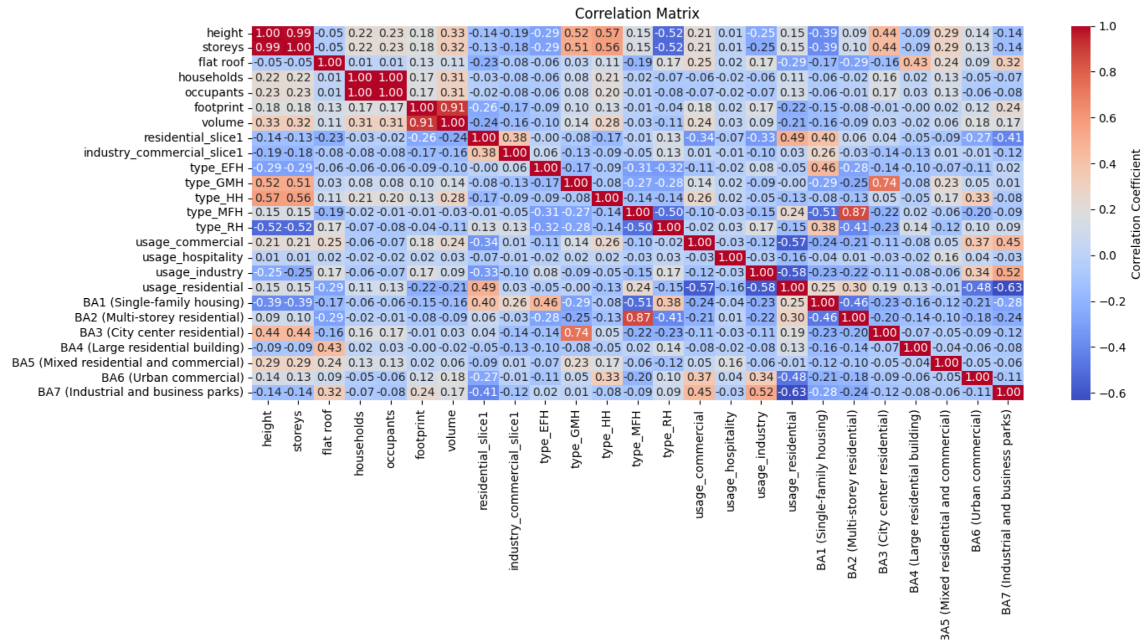

Figure 3. Correlation matrix showing correlation between the different parameters

(red)	BA1 (Detached Single-Family Housing)
(blue)	BA2 (Multi-storey residential)
(green)	BA3 (City center residential)
(purple)	BA4 (Large residential building)
(orange)	BA5 (Mixed residential and commercial)
(brown)	BA6 (Urban commercial)
(pink)	BA7 (Industrial and business parks)

Table 3. Legend for coloring the buildings according to their building archetype

was applied, yielding the building archetype classifications as indicated in the same color coding as for Figure 2. On the right of both Figures, Google Earth images for parts of the chosen neighborhoods are depicted to illustrate the plausibility of the classification. As can be seen, the model correctly assigns BA3 *City center residential*, to the highlighted building (and its direct neighbors) in Figure 4, and BA1 *(Detached) Single-family housing* to the neighborhood in Figure 5.

4. Discussion

As the comparisons in Figures 4 and 5 and the correlations in Figure 3 show, the trained model produces plausible classifications. It should be noted that the comparison with aerial imagery serves as an illustration only; a quantitative validation of the classification of the complete test dataset would require a considerably larger labeled dataset and is thus left to future work. Even though the amount of existing training and validation data is still limited, the results in table 1 show promising results and most building archetypes can be classified correctly, with the possible exception of BA5 *Mixed residential and com-*

Figure 4. City center (BA3) assignment of a building with Google Earth comparison

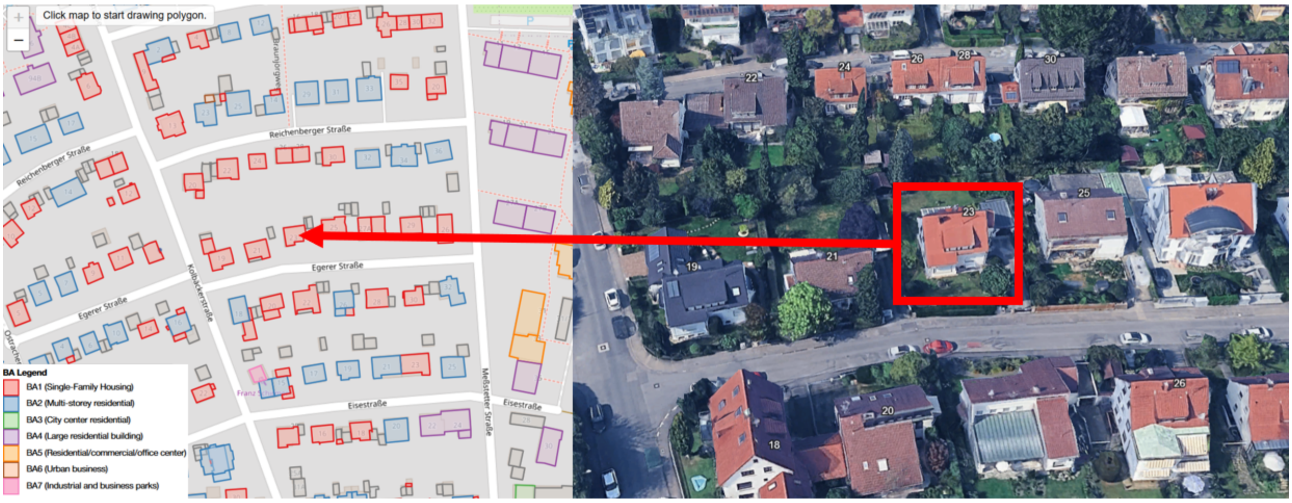

Figure 5. Detached single family housing (BA1) assignment of a building with Google Earth comparison

mercial, vs. BA6 *Urban commercial*. It has to be noted, though, that these are also the two archetypes that most closely resemble each other morphologically and based on surrounding-area information only.

The proposed method has the benefits of (i) being sector-agnostic, since the annotated data used for training as specified in section 2.2 includes general building-related information that is not linked to any specific industry or sector, and (ii) requiring only a standard CityGML file of the area of interest on top of the trained model, making it thus versatile and portable. The further derivation of the input parameters relies on SimStadt and, for the household and occupancy estimates, on aggregated German census statistics, both during training and during classification; however, no manual labeling, socio-economic microdata or further information per building beyond the CityGML file itself is required as input after model training.

As generally with AI-based model development, especially if the trained data is simulated data and not a measured ground truth, inconsistencies might arise from imperfect data. For example, some of the input data used is based on statistical information and thus unlikely to be correct on an individual building level; an example of this inaccuracy is the misclassification of some buildings as BA5 vs. BA6, as mentioned above (see also table 2). Inaccuracies might also stem from the still limited amount of training data, which leaves just 22 buildings for validation. Furthermore, CityGML files can contain errors that influence the output: next to geometric errors, buildings might be missing or might look different if they were (re-)built recently or can have incorrectly assigned usage codes. Lastly, if parts of a city quarter are selected from a larger data set, buildings at the edge of the selection have less surrounding buildings, potentially skewing the archetype classification due to a lack of information. To circumvent this, it might be useful to include a "buffer zone" in the data selection process to achieve more reliable results.

5. Conclusion

In sum, the proposed process and the trained model assigns archetypes to individual buildings based on CityGML-inherent properties per individual building and its surroundings with promising accuracy, given the still limited amount of training and validation data. The proposed method and process thus provides a tool to efficiently break down the built environment of cities into archetypical building types with very limited input data, allowing statistical, simulated or measured data to be annotated more precisely and, by this, enabling more granular and specific urban simulations. Since the proposed building archetypes, the process and the model used for model training are sector-agnostic, the method can be applied in a wide range of research areas. Prospective future applications span from research that, e.g., links household income levels and building archetypes, to more data-driven forms of municipal urban planning by investigating, e.g. linking various mobility ownership models to building archetypes and adapting planning practices.

Going forward, the confidence in the current model's precision could be improved by extending the training and validation data set. We are also considering merging BA5 and BA6 to simply represent commercial buildings within city quarters, which would alleviate the difficulties encountered in differentiating the two categories. Beyond that, testing and refining the model's validity for it to be also applicable beyond Germany might be a further next step.

References

Bao, K., Padsala, R., Thrän, D., Schröter, B., 2020. Urban Water Demand Simulation in Residential and Non-Residential Buildings Based on a CityGML Data Model. *International Journal of Geo-Information*, 9.

Dettmar, J., Drebes, C., Sieber, S., 2020. *Energetische Stadtraumtypen Strukturelle und energetische Kennwerte von Stadträumen*. 2, Fraunhofer IRB Verlag, Stuttgart.

Enteria, N., Santamouris, M., Eicker, U., 2021. *Urban Heat Island (UHI) Mitigation*. Springer.

Gulli, A., Pal, S., 2017. *Deep learning with Keras*. Packt Publishing Ltd.

Köhler, S., 2019. Stochastic generation of household electricity load profiles in 15-minute resolution on building level for whole city quarters. *Proceedings of the 16th IAEE European Conference: Energy Challenges for the Next Decade*, International Association for Energy Economics, Ljubljana, Slovenia.

Köhler, S., Betz, M., Duminil, E., Eicker, U., Schröter, B., 2021. A holistic approach to model electricity loads in cities. *Sustainability Management Forum— NachhaltigkeitsManagementForum*, 29number 2, Springer, 143–152.

Landesämter, S., 2022. Veröffentlichung — zensus2022.de. `https://www.zensus2022.de/DE/Ergebnisse-des-Zensus/02-veroeffentlichung.html`. [Accessed 23-04-2025].

Li, Y., Schubert, S., Kropp, J. P., Rybski, D., 2020. On the influence of density and morphology on the Urban Heat Island intensity. *Nature Communications*, 11(1), 2647. https://doi.org/10.1038/s41467-020-16461-9.

Nouvel, R., Brassel, K.-H., Bruse, M., Duminil, E., Coors, V., Eicker, U., Robinson, D., 2015. Simstadt, a new workflow-driven urban energy simulation platform for citygml city models. *Proceedings of International Conference CISBAT 2015 Future Buildings and Districts Sustainability from Nano to Urban Scale*, LESO-PB, EPFL, 889–894.

Nouvel, R., Zirak, M., Dastageeri, H., Coors, V., Eicker, U., 2014. Urban energy analysis based on 3d city model for national scale applications. *IBPSA Germany conference*, 8, 83–90.

Otto, R., Singer, J., von Puttkamer, B., Louis, D., Casey, L., Duminil, E., Schröter, B., 2026. Projected ramp-up of heat pumps, rooftop PV, and BEVs in Germany to 2045 across neighborhood archetypes. *Energy Reports*, 15, 109118. https://www.sciencedirect.com/science/article/pii/S2352484726000879.

Rodríguez, L., Duminil, E., Ramos, J., Eicker, U., 2017. Assessment of the photovoltaic potential at urban level based on 3D city models: A case study and new methodological approach. *Solar Energy*, 146, 264–275.

Singh, P., Manure, A., Singh, P., Manure, A., 2020. Introduction to tensorflow 2.0. *Learn TensorFlow 2.0: Implement Machine Learning and Deep Learning Models with Python*, 1–24.

Theile, P., Kesnar, C., Czock, B. H., Moritz, M., Novirdoust, A. A., Coors, V., Wagner, J., Schröter, B., 2022. There's no place like home – The impact of residential heterogeneity on bottom-up energy system modeling. *Energy and Buildings*, 254, 111591. https://www.sciencedirect.com/science/article/pii/S0378778821008756.

van Rijsbergen, C. J., 1979. *Information Retrieval*. Butterworth-Heinemann, London.

Weiler, V., Lust, D., Brennenstuhl, M., Brassel, K.-H., Duminil, E., Eicker, U., 2022. Automatic dimensioning of energy system components for building cluster simulation. *Applied Energy*, 313, 118651.

Zeynali, R., Mandanici, E., Bitelli, G., 2025. Urban Heat Island Dynamics in Bologna: Assessing the Influence of Urban Morphology on Air and Surface Temperatures.

Zhu, M., Jin, J., 2025. Data-Driven Urban Digital Twins and Critical Infrastructure Under Climate Change: A Review of Frameworks and Applications. *Urban Planning*, 10(0). https://www.cogitatiopress.com/urbanplanning/article/view/10109.