AUTOMATED BUILDING CLASSIFICATION FOR SCALABLE URBAN MATERIAL STOCK ASSESSMENT

Elisa Iliste, Ingrid Viskus, Ergo Pikas and Targo Kalamees
Tallinn University of Technology, Tallinn, Estonia

Abstract

Soviet-era apartment blocks in Eastern Europe are being demolished, generating mixed waste and limiting circular reuse. Existing assessments rely on generic archetypes and lack element-level material detail. This study develops an automated algorithm to detect standardised Soviet building series for scalable material stock analysis. A distance-based classification using apartment layouts, floor area, height, and staircases was validated on 337 buildings and applied to Estonia's building stock. The model achieved 70% series-level accuracy, with dominant series represent 61% of the stock. This element-level quantification enables targeted deconstruction and higher material recovery for circular economy planning.

Introduction

Across Eastern Europe, a substantial share of the housing stock consists of Soviet-era buildings constructed during periods of mass housing development. Built under conditions prioritising speed and quantity over long-term durability (Meuser and Zadorin, 2015), now exhibiting declining living conditions, deteriorating structural integrity, and architectural obsolescence.

The Soviet-era housing stock in Estonia is characterised by residential districts dominated by standardised pre-fabricated concrete apartment buildings. These structures, ranging primarily from 3- 5 to 9 and 16-storey tower blocks, were assembled using a limited number of industrial mass-produced series (e.g 1-317, 1-464, 111-121, etc.). Visually, these districts present a repetitive landscape of rectangular blocks grouped around semi-open courtyards or distributed freely to ensure better daylight conditions (Glendinning, 2019).

Figure 1: Example of a typical 5-storey 111-121 series building

Many of the soviet era buildings are facing abandonment; they no longer function as viable housing, both socially and economically, leading to deterioration and inevitable demolition (Eurofound, 2023). Traditional demolition practices generate large quantities of mixed construction and demolition waste, much of which is directed to landfilling or low-value processing, reinforcing a linear "produce-use-discard" economic model (Ilomets *et al.*, 2022). However, research highlights that demolition waste contains substantial amounts of materials—such as concrete, masonry, and timber—that have high reuse or recycling potential if handled appropriately (Sakthibala *et al.*, 2025). Selective deconstruction, as opposed to conventional demolition, enables materials to be separated at source, significantly improving their quality and recyclability while retaining economic value (Ilomets *et al.*, 2022). Timeliness is critical: early intervention allows material recovery before deterioration, maximising both environmental and economic benefits.

Despite this circular economic opportunity, current approaches to assessing urban material stocks face significant limitations. Research on material potential in existing buildings is constrained by data availability (Romero Perez de Tudela *et al.*, 2020; Tirado *et al.*, 2021). Urban analytics typically rely on building archetypes based on usage type and construction year only, without differentiating material composition, many rely only of building volume or area (Ali *et al.*, 2021; Oliveira *et al.*, 2024). In Urban Energy Modelling, for example, LOD1 or LOD2 models are used to calculate building volumes (Xu *et al.*, 2023), making them appropriate for envelope material quantities but failing to represent building interiors. From machine-learning applications (Markopoulou *et al.*, 2024) used image-based segmentation with cadastral data, and (Ghione *et al.*, 2022) used CNN-based building classification for Oslo's building stock. The main limitation of these image-based solutions is the computational efficiency.

Soviet-era architecture offers a unique opportunity: its centralised design templates and documented material specifications make it an ideal candidate for systematic detection and material quantification. Many of these buildings are remotely located, built for factory workers, and thus disproportionately affected by population decline (Glendinning, 2019). While collections of original design drawings exist (Meuser and Zadorin, 2015). Much of the documentation is largely available only in Russian,

limiting accessibility. Recent advances in AI-powered translation and computer vision now make these previously unviable data sources accessible for large-scale analysis.
If building series—can be reliably detected from readily available building data, it becomes possible to estimate material stocks at urban and national scales without costly individual building surveys. This is essential for circular economy planning, enabling municipalities to identify buildings with high reuse potential, prioritise selective deconstruction over conventional demolition, and plan material recovery operations before building deterioration advances.
This research addresses this opportunity by developing an automated detection algorithm for building series using open building registry data. The design science research methodology is employed to develop and evaluate the detection algorithm and demonstrate its application for material stock quantification. The algorithm is designed to operate with minimal input parameters, while maintaining sufficient accuracy for practical decision-making. To operationalise the main aim, the following research questions are posed:
RQ1: Can building series be automatically detected from open building registry data, and what classification accuracy can be achieved at the building stock level?
RQ2: Does this building series classification provide sufficient granularity to support circular economy planning and material stock quantification at element level?
This research makes three key contributions. First, it presents a novel distance-based classification approach for detecting building series using only four parameters (apartment configuration, floor areas, building height, and number of staircases). Second, it validates the method on Estonia's complete apartment building stock (13,200+ buildings), demonstrating up to 80% accuracy. Third, it demonstrates element-level material quantification for the 111-121 building series, revealing approximately 4.86 million m³ of concrete and 0.23 million tons of steel.

Methods

The design research methodology was employed to develop the building series and type detection algorithm. The development process comprised the following steps:

• Development criteria were defined
• Input parameters were selected based on data availability, relevance, and the objective of minimizing the number of required variables to ensure scalability and the possibility of manual data entry.
• Algorithm logic and calculation procedures were developed, including classification rules and parameter threshold definitions.
• The algorithm was implemented, tested and applied to the full Estonian apartment building dataset
• Validation was conducted using a smaller reference dataset to assess classification accuracy.
• Material quantities were calculated for one representative building series to demonstrate the applicability of the developed approach.

Data sets

A dataset of 437 buildings with expert-assigned series and types was utilised as ground truth. Out of those buildings, 65% had information about both the series and the extension. The other 35% of buildings were untypical, meaning they didn't have a series assigned to them. This set was divided into two. First, a set of 337 apartment buildings was used to validate the workflow itself. Another 100 buildings were used for additional validation for the national-level mapping.
Additionally, a dataset containing all Estonian apartment buildings was utilised for the nation-level mapping. Before applying the workflow, the dataset was pre-processed. Wooden apartment buildings (approximately 5,700) were excluded because the template library does not include wooden buildings. Buildings constructed before 1945 (approximately 1,300 buildings) were also excluded, as the template set is limited to post-war standardized designs, corresponding to the established classification of Estonian apartment building archetypes. After this filtering 13,200 buildings remained.

Data Sources

The Estonian Building Register (EBR) is a public database that contains information on all Estonian buildings. It includes various types of data, such as the year of first use, purpose of use, dimensions (footprint area, heated area, and building height), apartment-specific data (apartment number, floor of entry, area, number of rooms) and building technical data that provide information on building envelope materials. It has an open data policy and web services for data acquisition.
A collection containing general information on more than 100 Soviet-era building designs was used to develop a database of templates serving as benchmarks for the distance-based classification. The level of detail and machine-readable data varied across the designs. Only those designs containing the parameters required for the detection algorithm were included in the final template database.

Evaluation Methods

Recall was used to measure the proportion of correctly identified buildings for each series and is defined as:

$$\text{Recall}_s = \frac{TP_s}{TP_s + FN_s} \qquad (1)$$

where TP_s is the number of true positives for series s, and FN_s is the number of false negatives (i.e. buildings belonging to series s that were assigned to a different series). High recall indicates that the method is effective at detecting buildings of a given series when they are present.
The overall classification error rate is defined as the proportion of incorrectly classified buildings:

$$\text{Error rate} = \frac{N_{\text{incorrect}}}{N_{\text{total}}} \qquad (2)$$

where $N_{\text{incorrect}}$ is the number of buildings whose predicted series does not match the validated series, and N_{total} is the total number of evaluated buildings. In addition, a filtered error rate was computed by excluding buildings that do not correspond to any known template ("untypical" buildings), providing an estimate of performance for standardised building types only.

The false-positive rate for a given series measures the extent to which buildings from other series are incorrectly assigned to it:

$$\text{FP rate}_s = \frac{FP_s}{TP_s + FP_s} \quad \textbf{(3)}$$

where FP_s is the number of buildings incorrectly assigned to series s. High false-positive rates indicate that a series acts as an attractor for structurally similar but distinct building series.

Building Series and Type Representation

General data from 75 designs across 11 series were compiled into design templates to serve as benchmarks for distance calculations. Templates were implemented as Python objects using a set of data classes. Each template encodes one building design and is described by six main attributes, sampled in Table 1.

First, the series and series extension (type) identify the building family and the specific design variant (e.g., 1-464-1KE). Together, they form a unique series identifier.

Second, the series type indicates whether the template represents a whole building (WHOLE) or a modular staircase or block (MODULE). The whole series corresponds to complete buildings with multiple staircases, whereas the modular series describes a single repeatable building staircase.

Third, the spatial pattern describes the apartment configuration on a typical floor. This is represented as a nested list. For whole buildings, each inner list corresponds to one staircase and contains the number of apartments in each segment of that staircase. For modular buildings, the pattern consists of a single list representing one module.

Fourth, the area pattern stores the floor area associated with each staircase. Instead of storing the area of individual apartments, the total area per staircase is used.

Fifth, the number of floors is represented as a list. For buildings with a fixed height, this list contains a single value. For building series occurring at multiple heights, the list includes all typical numbers of floors.

Finally, all templates are stored in a central template library that serves as a registry and allows filtering by series type or other attributes.

This representation provides a machine-readable encoding of architectural typologies, while preserving the hierarchical structure of real buildings.

Table 1: Template instance examples, Value A represents a modular type, Value B represents the whole building type

Attribute	Value A	Value B
Series	'111-121'	'1-464'
Extension	'03E'	'17'
Type	Module	Whole
Pattern	[[4, 4, 2, 3]]	[[2,1,3], [2,2,2], [2,2,2], [3,1,2]]
Area Pattern	[255.0]	[136, 136, 136, 136]
Floor nr	[9]	[4]

Developed Workflow

The developed building series detection workflow consists of three main phases: (1) Data Augmentation and Preprocessing; (2) Distance calculation; (3) Classification and Scoring. The workflow is illustrated on Figure 2.

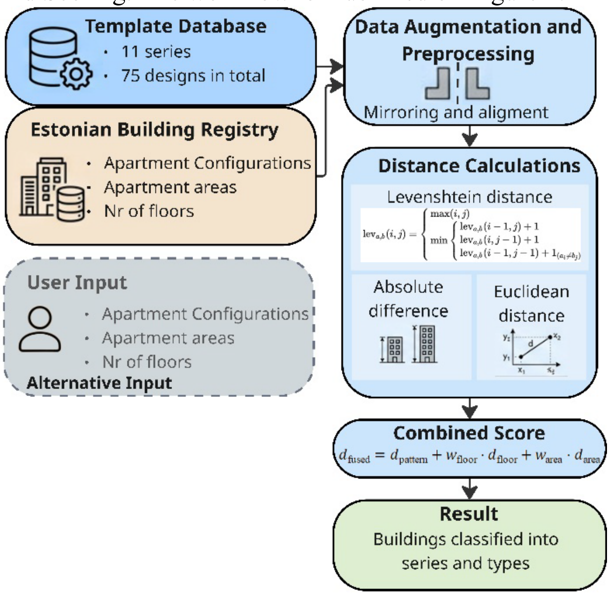

Figure 2: Series detection workflow

Three principal design criteria guided the development of the framework:

• Machine-readable and scalable representation of the building series, the data must be easily expandable with new information

• Robustness to structural variation and data inconsistencies:

Data Augmentation and Pre-Processing

To account for symmetric building layouts and staircase ordering inconsistencies, the matching procedure evaluates both the original and mirrored configurations of each building. Mirroring is implemented by reversing the order of staircases in the building representation. This accounts for cases where the same building design may be recorded with opposite orientations in the source data. For each building, two configurations are therefore considered: the original staircase sequence and its mirrored counterpart. For each configuration, the full matching workflow is executed independently.

Modular building series have to be treated differently from whole-building series. While whole-building templates represent complete buildings with a fixed configuration, modular templates represent configurable staircase units that may appear multiple times within a building and be mixed.

For modular series, templates are first grouped by their base series code. Each staircase of the subject building is then independently matched against all modular templates within the corresponding series. The best matching module for each staircase is selected using the minimum Levenshtein distance, considering both the original and mirrored module patterns.

The total distance for a modular series is computed as the sum of the per-staircase distances:

$$d_{\text{module}} = \sum_{i=1}^{n} \min\left(dL(s_i, m_j)\right) \quad \textbf{(4)}$$

where s_i is the apartment pattern of a staircase i in the building, and m_j are the available module patterns for the series. For each modular series, only the configuration with the smallest distance is selected for further calculations to minimise computational cost.

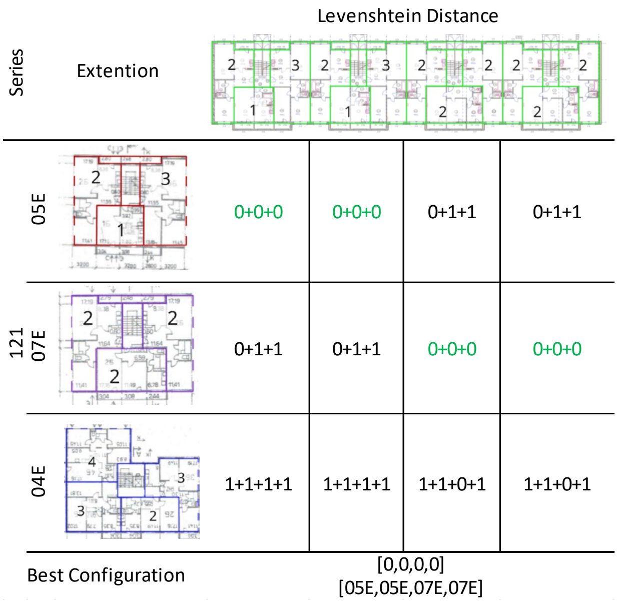

Figure 3: Example of finding the best modular configuration over a series using the Levenshtein distance

Distance Calculations

Apartment Pattern distance

Pattern similarity is computed using the Levenshtein (edit) distance between the subject building's staircase sequences and the template sequences. This metric was chosen because the objective is to identify the most similar structural configuration rather than to measure numerical proximity. The staircase sequences represent categorical configurations rather than continuous numerical variables; therefore, computing absolute or Euclidean differences would not meaningfully capture the similarity. The Levenshtein distance is calculated as shown in Equation 5.

$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j) + 1 \\ \operatorname{lev}_{a,b}(i,j-1) + 1 \\ \operatorname{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \quad \textbf{(5)}$$

The algorithm considers both original and mirrored building configurations.

Floor distance

The floor distance quantifies the difference between building floor count and the template specifications:

$$d_{\text{floor}} = \min_{f \in F} \mid n_{\text{floors}} - f \mid \quad \textbf{(6)}$$

where f is the set of allowed floor counts from the template and $\boldsymbol{n}_{\text{floors}}$ is the actual building floor count.

Area Distance

Area similarity is calculated using the Euclidean distance, with each staircase acting as a dimension. For building staircase areas $\mathbf{A}_B = [a_1, a_2, \dots, a_n]$ and template areas $\mathbf{A}_T = [t_1, t_2, \dots, t_n]$:

$$d_{\text{area}} = \sqrt{\sum_{i=1}^{n} \delta_i^2} \quad \textbf{(7)}$$

where the percentage difference for each valid staircase pair is:

$$\delta_i = \begin{cases} \min\left(\dfrac{\mid a_i - t_i \mid}{t_i}, 1.0\right) & \text{if } a_i > 0 \text{ and } t_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(8)}$$

Fused Distance Score

All distances are combined using a weighted linear combination:

$$d_{fused} = d_{pattern} + w_{floor} \cdot d_{floor} + w_{area} \cdot d_{area} \quad \textbf{(9)}$$

where w_{floor} and w_{area} control the relative importance of floor and area constraints and allow to prioritize the pattern distance. In this paper, these were chosen as 0.5.

Validation of the workflow

337 buildings were used to assess the performance of the proposed workflow. The overall classification accuracy was 60% for series and 50% for templates. When excluding the generic unclassified category, accuracy increases to 80% for series and 70% for templates, indicating a higher performance for well-defined building types. To assess scalability, the runtime of the classification workflow was measured.

The violin plot shows that while the average distance for correct predictions is 1.2, the median distance is lower (0.4), indicating that the majority of correctly classified buildings have very small distances, with the mean being influenced by a limited number of high-distance outliers (mainly the unassigned buildings, which have an average distance of 3.5). The average and median distances for incorrectly assigned templates are 2.8 and 2.7, respectively.

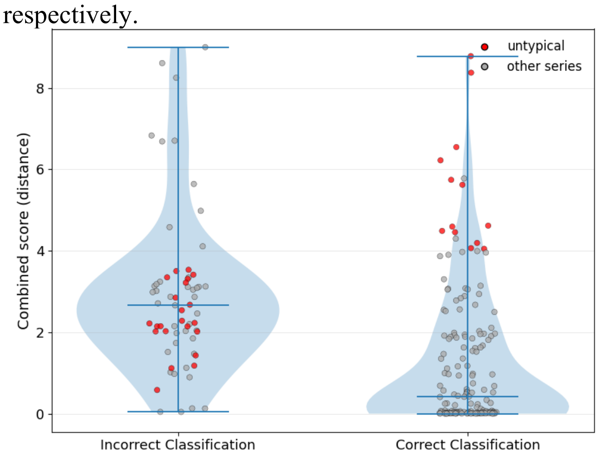

Figure 4: Distribution of the combined distance score by correctness, showing separation between correctly and incorrectly classified buildings.

To assess the performance of the proposed classification approach, we analysed the confusion matrix(Table 2) between the validated building series and the final predicted series. Columns represent the ground-truth (validated) series, while rows represent the model predictions.

The model demonstrates strong performance for several major building series. High, over 90% recall is observed for 121(100%),72057 (100%),10(93%),133(92%) and 1-464 (91%). These results suggest that the proposed features effectively capture the characteristics of the most well-defined building types.

In contrast, several series exhibit poor recall. Series 113-EKE, 66, 22 and 3F-2F all have a recall of 0. indicating that these categories are either underrepresented in the training data or insufficiently distinguishable from other series. The generic category also performs poorly, correctly classifying only 37 % of true instances.

The errors show consistent confusion between specific series. For example,

- 1Э-318 → 121 (36%), indicating that over one third of buildings from series 1Э-318 are systematically misclassified as 121.
- 113-EKE 10 → 23 (50%), suggesting strong similarity between these two series.

These results indicate families of series that are structurally similar and difficult to distinguish using a flat multi-class classification scheme.

Several series show a disproportionate number of false positives. Series 121 absorbs 36% of true 1Э-318, 11% of true Other, and 3% of true 1-317, despite having perfect recall. Similarly, the Other class accounts for 18% of true 1э-318, 10% of true 113-EKE 11, and 7% of true 133. Series 23 also generates notable false positives, including 50% of true 113-EKE 10 and 13% of true 113-EKE 11.

Nation-level mapping

The combined distance metric was computed for the 13,200 buildings. The full building stock (13,200 buildings) was processed in approximately seconds on a standard desktop machine. As a result of this classification, the most frequent series were 1-317 (14.26%), 10 (13.47%), 111-121 (13.00%), 72057 (11.29%), and 23 (9.45%). Together, these five series account for over 61% of all classified buildings.

The assigned building series was validated using a sample of 100 buildings. The overall classification error rate was 36%; excluding buildings that did not correspond to any known template reduced it to 28%. The recall remained high for 1-464 (98%); 1-317 also showed a high recall of 82%.

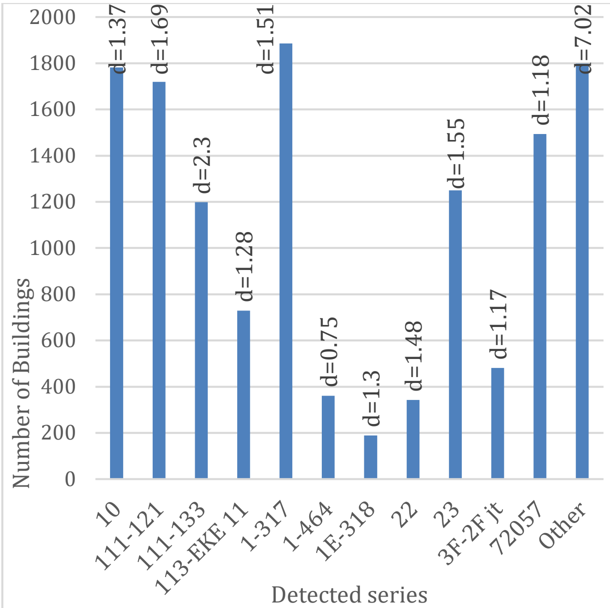

Figure 5: Building Stock Classification by Detected Series with Average Distance Score (d)

Material quantities calculation

To demonstrate the proposed workflow, a material stock calculation was performed for the 111-121 series. The series was selected for two reasons: firstly, it is one of the most prevalent building types in the national stock

Table 2: Confusion Matrix, in the column are the validated series, in the row the automatically classified series, green indicates a match and red the correct classification

Result / Validated	10	22	23	66	121	133	72057	113-EKE 10	113-EKE 11	113-EKE 12	1-317	1-464	1Э-318	Other	3F-2F
10	94%	50%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	6%	0%
22	0%	0%	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	33%
23	6%	0%	69%	0%	0%	0%	0%	50%	13%	0%	0%	8%	9%	3%	0%
121	0%	0%	0%	100%	100%	0%	0%	0%	0%	0%	3%	0%	36%	11%	0%
133	0%	50%	0%	0%	0%	93%	0%	0%	0%	100%	0%	0%	0%	11%	0%
72057	0%	0%	0%	0%	0%	0%	100%	0%	3%	0%	3%	0%	0%	14%	33%
113-EKE	0%	0%	17%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
113-EKE 11	0%	0%	0%	0%	0%	0%	0%	50%	71%	0%	0%	0%	0%	0%	0%
1-317	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	87%	0%	0%	11%	0%
1-464	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	92%	0%	0%	0%
1Э-318	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	36%	0%	0%
3F-2F	0%	0%	6%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%	6%	33%
Other	0%	0%	3%	0%	0%	7%	0%	0%	10%	0%	8%	0%	18%	37%	0%

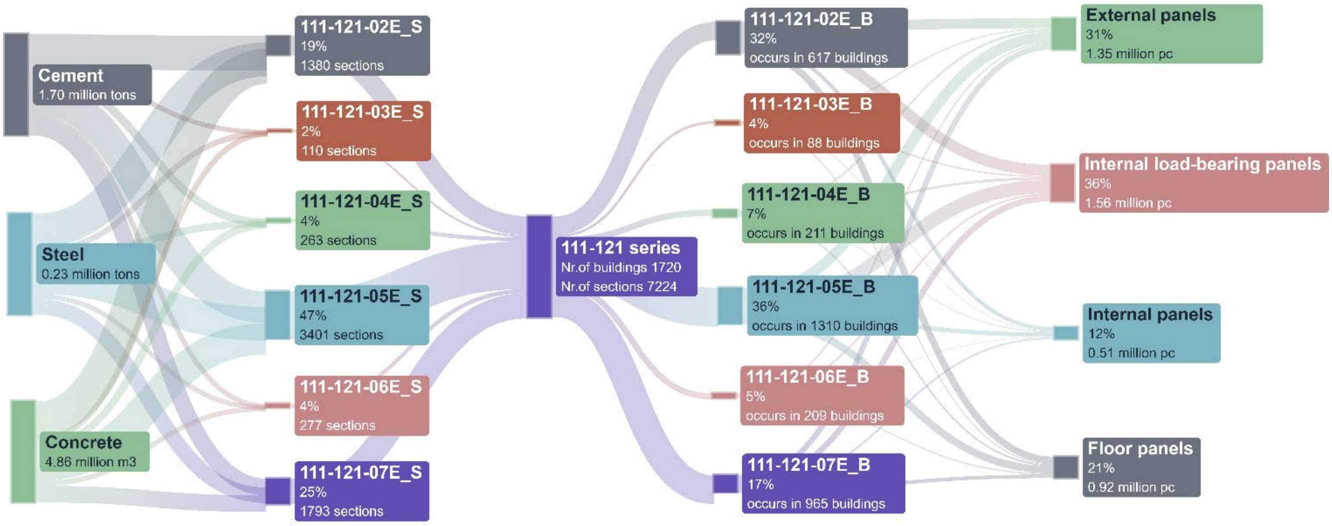

Figure 6: Material flow analysis for the 111-121 series: Central nodes represent building section templates, with flows splitting into two paths: left quantifies raw materials (concrete, steel, cement) for recycling estimates, and right classifies prefabricated components (floor and wall panels) for reuse planning.

(N=1720 detected buildings); secondly, the 111-121 series is modular, consisting of various section types assembled in different configurations. This modularity allows analysis in which material quantities are calculated for a specific section template rather than using generic building averages.

The analysis focused on the six most frequently detected types in the series, which collectively account for the vast majority of the 111-121 building stock.

Total material stock (M*total*) for the series was calculated by aggregating the specific material quantities of each identified section type:

$$M_{total} = \sum_{i=1}^{k} \left(N_i \times Q_{(i,m)} \right) \tag{10}$$

Where:

Ni is the total count of detected sections of type *i* across the entire building stock (derived from the dataset's frequency distribution; e.g., type 111-121-05E appears 3401 times).

Qi,*m* is the quantity of a specific material *m* (e.g., reinforced concrete in *m*3, steel in tons) per single section of type *i*, extracted from the original design passports (reference project documentation).

k is the number of unique section templates identified (e.g., 02E, 05E, 07E, etc.)

As shown in Figure 6 The material flow analysis reveals distinct patterns. Although 5-storey building sections (Types 05E and 07E) are numerically dominant, accounting for over 70% of all sections, the 9-storey sections contribute disproportionately to the total material mass due to their larger building volume. The analysis indicates a total embodied stock for series 111-121 of approximately 4.86 million m3 of concrete, 1.70 tons of cement, and 0.23 million tons of steel for this series alone. Breaking this down by element type, the floor slabs (panels) account for the largest share at ~36%. This level of granularity is crucial for circular planning as it distinguishes between elements based on their material complexity and 'cleanness', which are the key factors in evaluating the reuse potential. Unlike the external wall panels, which in Soviet-era construction often consist of composite layers that are difficult to separate and whose insulating layers are often problematic to reuse, floor and internal load-bearing panels are typically monolithic reinforced concrete elements. That according to (Ilomets *et al.*, 2022) This structural homogeneity gives floor panels a higher potential for direct reuse, with an estimated recovery of up to 90% for floor slabs in good condition. He also confirms that internal load-bearing elements are one of the most valuable components, as they are generally well-preserved and clean.

Discussion

The results demonstrate that building series can be systematically identified from open building registry data with meaningful accuracy. For the primary validation dataset (337 buildings), the overall classification accuracy was 60% for series and 50% for templates. When excluding buildings that do not correspond to any known template (generic/unclassified category), the accuracy increases to 80% for series and 70% for templates. At the national level, validation yielded an overall accuracy of 64%, which rose to 72% when excluding unclassified buildings.

This performance is comparable to machine learning approaches reported in the literature, such as the 87% accuracy reported by Markopoulou et al. (2024). However, the distance-based approach offers distinct advantages in interpretability, minimal data requirements, and computational efficiency. Previous material stock studies have relied on statistical averages, applying material intensity coefficients (*kg*/*m*2 *or kg*/*m*3) to floor area or building volume (Oliveira *et al.*, 2024). While effective for large-scale estimates, these methods fail to capture the structural heterogeneity within building series. For example, taller sections require more reinforcement steel and concrete per m2 than lower counterparts due to structural requirements. The ability to detect these patterns enables more accurate quantification of material properties for engineering design rather than relying on statistical averages.

Although distance-based similarity metrics perform well across several building series, the results suggest that additional descriptive dimensions are needed to improve robustness, particularly for identifying atypical buildings that do not conform to any standardised series. Such cases are inherently difficult to capture using purely rule-based or metric-based approaches.

Machine learning, therefore, constitutes a promising direction for future work. While the current number of templates is too large to support a classification model with the available data, machine learning could be employed in a hybrid framework. For instance, a predictive model could assist in ranking candidate series rather than series and extensions, resolve ambiguities when multiple templates exhibit similar distance values, and flag buildings that deviate substantially from known typologies.

The implications of the results extend beyond quantifying material stock. Building series detection enables the systematic application of any series-specific data at the urban scale. This includes thermal properties and spatial data for energy modelling, as well as structural characteristics and construction details for renovation planning.

While the proposed method is developed and demonstrated on standardised Soviet-era building typologies, the underlying framework is not unique to this context. Similar industrialised and prefabricated housing approaches were widely implemented across Europe during the post-war period. These systems share key characteristics with Soviet building series, including modular design, limited layout variation, and predefined structural and spatial configurations.

Comparable patterns can also be observed in other European contexts, such as the Swedish Million Programme (Berggren and Wall, 2019), where large volumes of residential buildings were constructed using standardised designs. Although the specific notion of “series” may not be explicitly defined in all these contexts, the underlying logic of repetition is consistent.

Several limitations remain. First, data availability and quality significantly affect applicability and scalability. While Estonia's comprehensive building registry enabled national-scale analysis, many Eastern European countries lack a comparable open data infrastructure.

Second, parameter sensitivity poses a critical challenge. Building registries often contain inconsistencies and data quality issues that directly affect large-scale assessments that rely on these sources. Errors in registry data propagate through the classification workflow, leading to potential misclassification of building series. Consequently, the classification accuracy is highly sensitive to the quality of input parameters. Small deviations in recorded apartment configurations, floor areas, or building heights can lead to incorrect template assignments. The confusion matrix revealed systematic misclassifications between specific series pairs (e.g., 1Э-318 → 121 at 36%, 113-EKE 10 → 23 at 50%), indicating that the four-parameter representation may be insufficient to distinguish closely related designs.

Third, template database coverage constrains comprehensiveness. The 75 designs represent major series but cannot capture the full diversity of Soviet-era construction. The "generic" category had only 37% recall, serving as a catch-all for buildings that did not match known templates. Expanding the database requires access to original design documentation.

Fourth, building modifications and renovations significantly impact classification accuracy. Soviet-era buildings have undergone various interventions during their 40-70-year lifespans: apartment reconfigurations that alter spatial patterns, balcony enclosures that change floor area, and roof additions that modify building height. The building registry may not consistently reflect these changes, leading to mismatches between recorded parameters and actual characteristics.

Despite these limitations, the method provides a practical foundation for circular economy planning where detailed building-level surveys are infeasible. The element-level material quantification demonstrated for the 111-121 series offers actionable information for selective deconstruction planning. Future research should expand the validation dataset, focus on hybrid classification frameworks that combine distance metrics with machine learning, conduct validation studies in other Eastern European countries, and integrate building condition assessment to account for deterioration and modifications.

Conclusions

This paper proposes an automated series of building detection for urban material stock assessment, focusing on Soviet-era standardised buildings in Estonia.

The results show that building series can be automatically detected from open building registry data with meaningful accuracy. Using four parameters: apartment configuration, floor areas, building height, and number of staircases, the distance-based algorithm achieved 60% overall accuracy for series and 50% for templates across 337 validation buildings. Excluding the generic unclassified category, accuracy increased to 80% for series and 70% for templates. National-scale classification of 13,200 buildings yielded a similar accuracy of 64% for all series, increasing to 72% when excluding non-standardised buildings.

Building series classification provides sufficient granularity for circular economy planning and element-level material stock quantification. The 111-121 series (1,720 buildings) revealed approximately 4.86 million m³ of concrete, 0.23 million tons of steel, and 1.70 million tons of cement. Floor slabs account for 36% of total material mass. This granularity is essential because it distinguishes elements by reuse potential—monolithic floor panels offer up to 90% recovery, compared with composite external wall panels with problematic insulation layers.

This work provides three key contributions: (1) Novel distance-based classification with minimal input parameters maintaining practical accuracy; (2) Validation on Estonia's complete apartment building stock, demonstrating national-scale scalability; (3) Element-level material quantification, distinguishing components by reuse potential.

Acknowledgments

This work has been supported by the Estonian Research Council grants (PSG963 and PRG2732); by the European Commission through Horizon Europe projects SoftAcademy (EUI01- 168); and by the Estonian Centre of Excellence in Energy Efficiency, ENER (grant TK230) funded by the Estonian Ministry of Education and Research.

References

Ali, U., Shamsi, M.H., Hoare, C., Mangina, E. and O'Donnell, J. (2021), "Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis", *Energy and Buildings*, Vol. 246, p. 111073, doi: 10.1016/j.enbuild.2021.111073.

Berggren, B. and Wall, M. (2019), "Review of Constructions and Materials Used in Swedish Residential Buildings during the Post-War Peak of Production", *Buildings*, Multidisciplinary Digital Publishing Institute, Vol. 9 No. 4, doi: 10.3390/buildings9040099.

Eurofound. (2023), *Unaffordable and Inadequate Housing in Europe*, Publications Office of the European Union, Luxembourg.

Ghione, F., Mæland, S., Meslem, A. and Oye, V. (2022), "Building Stock Classification Using Machine Learning: A Case Study for Oslo, Norway", *Frontiers in Earth Science*, Frontiers, Vol. 10, doi: 10.3389/feart.2022.886145.

Glendinning, M. (2019), "Mass housing and extensive urbanism in the Baltic countries and Central/Eastern Europe: a comparative overview", *Housing Estates in the Baltic Countries: The Legacy of Central Planning in Estonia, Latvia and Lithuania*, Springer International Publishing Cham, pp. 117–136.

Ilomets, S., Kiviste, M., Tuisk, T., Paalandi, K., Kallavus, U., Hain, T., Pikas, E., *et al.* (2022), *Applied Study on Reuse and Recycling of Materials from Demolition of Vacant Apartment Buildings – Phase 1 Interim Report*, Ministry of Economic Affairs and Communications, Tallinn.

Markopoulou, A., Taut, O. and Shawqy, H. (2024), "Enhanced Databases on City's Building Material Stock. An Urban Mining Method Based on Machine Learning for Enabling Building's Materials Reuse Strategies", in Thomsen, M.R., Ratti, C. and Tamke, M. (Eds.), *Design for Rethinking Resources*, Springer International Publishing, Cham, pp. 685–701, doi: 10.1007/978-3-031-36554-6_44.

Meuser, P. and Zadorin, D. (2015), "Towards a typology of soviet mass housing", *Berlin: DOM. Alexander*.

Oliveira, J., Schreiber, D. and Jahno, V. (2024), "Circular Economy and Buildings as Material Banks in Mitigation of Environmental Impacts from Construction and Demolition Waste", *Sustainability*, Vol. 16, p. 5022, doi: 10.3390/su16125022.

Romero Perez de Tudela, A., Rose, C.M. and Stegemann, J.A. (2020), "Quantification of material stocks in existing buildings using secondary data—A case study for timber in a London Borough", *Resources, Conservation & Recycling: X*, Vol. 5, p. 100027, doi: 10.1016/j.rcrx.2019.100027.

Sakthibala, R.K., Vasanthi, P., Hariharasudhan, C. and Partheeban, P. (2025), "A critical review on recycling and reuse of construction and demolition waste materials", *Cleaner Waste Systems*, Vol. 12, p. 100375, doi: 10.1016/j.clwas.2025.100375.

Tirado, R., Aublet, A., Laurenceau, S., Thorel, M., Louërat, M. and Habert, G. (2021), "Component-Based Model for Building Material Stock and Waste-Flow Characterization: A Case in the Île-de-France Region", *Sustainability*, Vol. 13 No. 23, doi: 10.3390/su132313159.

Xu, H., Kim, J.I., Zhang, L. and Chen, J. (2023), "LOD2 for energy simulation (LOD2ES) for CityGML: A novel level of details model for IFC-based building features extraction and energy simulation", *Journal of Building Engineering*, p. 107715, doi: 10.1016/j.jobe.2023.107715.