# AUTOMATIC ROOF TYPE CLASSIFICATION FROM CITYJSON FOR RULE-BASED ROOF STRUCTURE'S MATERIAL QUANTITY ASSESSMENT

Dušan Isailović[1], Aleksandra Parezanović[1], Zorana Petojević[1,2], Nenad Brodić[1]
[1]University of Belgrade, Faculty of Civil Engineering, Serbia;
[2]ZHAW School of Architecture, Design, and Civil Engineering, Zurich, Switzerland

## Abstract

Current city-scale material stock assessment relies on archetype-based estimation with limited geometric grounding or on city-to-BIM conversion workflows that are costly and sensitive to representation detail. Both hinder auditable roof typology and quantity derivation from open 3D city models. This paper proposes an interpretable CityJSON pipeline that merges roof meshes into planar patches, extracts roof metrics (area, projection, slopes, span), performs rule-based roof classification (flat/shed/gable/hip/complex), and computes rule-based roof-layer and timber quantities. On 88 labelled buildings, LoD2 achieves 0.716 accuracy (macro-F1 0.543); BIM comparison on six buildings supports feasibility. The contribution is a reproducible, explainable roof-to-quantities method with quantified LoD sensitivity.

## Introduction

Cities are increasingly adopting open 3D city models to support planning, permitting, resilience, and environmental assessment. At the same time, circular construction and urban mining require reliable, scalable estimates of material stocks and flows in the existing building stock—especially for elements that dominate mass and replacement cycles. Roofs are a particularly relevant target: they strongly influence building performance and refurbishment needs, they exhibit recurring typologies that relate to structural systems, and they are well represented in city models through explicitly modeled roof surfaces. Yet turning city-model roof geometry into material quantities remains challenging because the data are typically triangulated, vary across levels of detail (LoD), and rarely contain construction semantics beyond coarse surface types.

Most existing city-scale quantity and stock assessment approaches rely on one of two strategies. The first is archetype-based estimation, where building age, use, and generic typologies are used to assign typical assemblies and materials. These methods can scale well, but their accuracy depends heavily on assumptions and weak geometric coupling. The second strategy is a geometry-driven pipeline that extracts measurements from 3D models, then maps them to construction rules. While promising, such pipelines must address practical issues: roof geometry is often fragmented into many small facets, different LoDs represent the same roof differently, and black-box methods (e.g., deep learning) can be difficult to audit when outputs directly drive normative structural selection and material take-offs.

This paper argues that, for city-scale roof material quantification, interpretability and robustness to representational variability are as important as raw classification accuracy. Instead of relying on end-to-end learning, we propose a deterministic workflow that explicitly links roof geometry to roof type labels and quantity-relevant metrics, enabling traceable propagation from input model to material quantities. The workflow uses CityJSON as an open and lightweight representation and focuses on (i) merging triangulated roof faces into planar roof patches, (ii) extracting roof metrics needed by rule-based quantity models, and (iii) classifying roof types using explainable geometric rules. Because LoD can substantially alter planar fragmentation and rule triggers, we explicitly evaluate the same classifier on both LoD2 and LoD3 representations of the same area to quantify typology stability and to support guidance on which LoD is more appropriate for operational quantity pipelines.

The method is demonstrated on a Belgrade test area with 88 manually labelled buildings. In addition, a subset of six buildings is reconstructed in BIM to provide baseline take-offs for validating rule-based roof-layer and timber quantities. Beyond reporting classification performance, we analyze how roof typology errors affect downstream material estimates—showing that some materials (e.g., timber) are relatively robust to subtype ambiguity, while others (notably concrete in flat-roof archetypes) are highly sensitive to flat/non-flat misclassification due to archetype switching.

## Related work

### 3D city models as a basis for urban-scale assessment

Open 3D city models (CityGML/CityJSON) are increasingly used as a geometric backbone for city-scale analyses because they provide consistent building envelopes and semantic surface classes (e.g., RoofSurface) that can be processed automatically. CityJSON in particular has been positioned as a lightweight, developer-friendly encoding of CityGML concepts for practical processing workflows (Ledoux, et al., 2019). The Levels of Detail (LoD2, LoD3) used in such models follow the conceptual definitions of CityGML 3.0, where LoD2 represents generalized roof structures and LoD3 introduces detailed architectural features such as dormers and roof superstructures (Open Geospatial Consortium, 2026).

Large-scale LoD2/LoD3 building models are commonly generated from airborne LiDAR and photogrammetric data using automated reconstruction pipelines that detect planar roof segments and infer roof topology (Haala & Brenner, 1999). These workflows often assign roof types explicitly as dataset attributes (e.g., *roofType* in CityGML), reflecting the inferred geometric archetype of the roof.

However, such attributes are not consistently available across datasets, are derived using dataset-specific assumptions, and are typically not designed for construction-oriented interpretation or material quantity estimation. As a result, relying directly on dataset-provided roof types may limit reproducibility and applicability across heterogeneous data sources.

**Roof type inference from city models and remote sensing**

Roof typology inference has been studied from multiple perspectives. A line of work targets roof shape classification for energy and solar potential assessment using machine learning on roof features, showing that roof-type prediction is feasible but sensitive to the quality and representativeness of training data and features (Mohajeri, et al., 2018). Other research infers roof type from sparse or derived building attributes, demonstrating that the “flat vs non-flat” decision can often be made robustly and that a small set of geometric attributes can be highly informative (Biljecki & Dehbi, 2019). Recent datasets and studies also highlight how roof typologies can be learned from CityJSON-labelled corpora, reporting roof classification performance in the 70–75% range across multiple European cities—comparable to the performance envelope observed in practical city-scale pipelines (Sitong, Chengzhi, Chi, & Qiuxian, 2023).

In parallel, deep learning approaches—often driven by aerial imagery—have been used to classify roof types (flat, gable, hip, etc.), benefiting from large training sets but typically offering limited traceability between the decision and the 3D geometry that will later drive construction quantities (Buyukdemircioglu, Can, & Kocaman, 2021). For applications such as material quantity estimation, where classification outcomes directly affect structural assumptions, interpretability becomes a critical requirement.

**BIM–GIS integration and material stock/quantity assessment**

A growing body of work addresses BIM–GIS integration for circular construction and city-scale material stock assessment. Recent studies combine remote sensing, 3D GIS and BIM to improve material estimation at building and element levels (Parezanović, Nadaždi, Isailović, Višnjevac, & Petojević, 2025). In the EC3 community specifically, approaches have also been proposed for converting 3D city models to BIM to support automated quantity take-offs (Dušan Isailović, Ana Nadaždi, Aleksandra Parezanović, Zorana Petojević, & Nenad Višnjevac, 2025). Complementary review work highlights interoperability challenges and the importance of consistent semantics in CityGML-based environments (Janisio-Pawłowska & Pawłowski, 2024).

At the urban scale, material stock datasets and frameworks commonly rely on city model geometry (footprints, heights, LoDs) combined with building attributes (age, typology, structural system) to compute material quantities (Pei, Biljecki, & Stouffs, 2022). While these approaches scale efficiently, they often rely on weak geometric coupling or require costly BIM conversion workflows.

**Positioning of this paper**

Across the literature, roof type inference is often treated as an intermediate step within 3D reconstruction pipelines or as a standalone classification task for applications such as solar potential assessment. In many operational datasets, roof types are already inferred during LiDAR- or photogrammetry-based reconstruction and may be available as attributes. However, these attributes are not consistently available, are dataset-specific, and are not designed for construction-oriented interpretation or material quantity estimation.

The novelty of this work lies not in roof classification alone, but in integrating roof plane extraction, geometric metric derivation, typology inference, and rule-based material quantity estimation into a single interpretable and auditable workflow. In this context, roof typology serves as a bridge between geometric representation and construction semantics, enabling traceable propagation from city-model geometry to material quantities.

Furthermore, this work explicitly investigates the sensitivity of roof typology inference to different Levels of Detail (LoD2 vs LoD3), highlighting how geometric representation affects both classification stability and downstream quantity estimation. This connection between representation, semantics, and quantities remains underexplored in existing literature.

## Method

The proposed workflow is an interpretable, city-scale pipeline that derives roof typology and quantity-relevant metrics from an open 3D city model. For each building, we:

1. Extract roof surfaces from CityJSON,
2. Merge triangulated roof faces into planar roof patches (“roof planes”),
3. Derive a roof-plane adjacency representation from shared edges,
4. Compute roof metrics required for material quantity rules,
5. Classify roof type using explicit geometric rules,
6. Map the classified roof type and span proxy to a normative structural archetype and compute timber quantities, and
7. Compute additional roof-layer quantities from roof area/projection using rule tables.

The workflow is executed primarily on LoD2; LoD3 is used to discuss representation-induced fragmentation effects and the resulting differences in typology stability.

The workflow uses:

- Cityjson lod2 (primary) and lod3 (comparative) building models for the Belgrade test area;
- Rule-based roof classifier implementation developed from the original roof_classifier.py into the lod2-selected version;
- Manual roof-type labels for evaluation; and
- Normative tables and rules for timber roof structure selection and quantity factors, plus roof-layer material rules.

The workflow produces:

- Roof type label (roof_type_automatic) and QA indicators;
- Roof metrics (roof area, horizontal projection area, slope statistics, plane count, span proxy);
- Normative structural type;
- Timber volume estimate;
- Roof-layer quantities; and
- For a BIM-modeled subset, a comparison against BIM-derived baseline take-offs.

For each building, roof faces are selected primarily using CityJSON semantics (e.g., RoofSurface), ensuring the method remains consistent with semantic organization where available. When semantics are incomplete, a geometric fallback can be applied by selecting upward-facing faces within the upper elevation band of the building envelope (reported as a limitation and used only when necessary).

**Planar roof patch (roof plane) extraction**

City models typically represent roofs as triangulated meshes; direct triangle-level reasoning is unstable and inflates the number of planes. Therefore, roof faces are merged into larger planar patches ("roof planes") using a deterministic clustering procedure:

1. Per-face geometry: for each face, compute a unit normal vector, face area, and a representative point (e.g., centroid).
2. Plane merging: faces are clustered into planes using (a) normal similarity (angle between normals below a tolerance) and (b) coplanarity (distance from the face representative point to the candidate plane below a tolerance). This normal + coplanarity merging prevents parallel-but-offset patches from being merged, which is important for detailed models where small roof features can create offset but parallel surfaces.
3. Plane attributes: for each roof plane, compute total area (sum of merged face areas), area-weighted mean normal, and slope angle relative to horizontal.

**Roof-plane adjacency representation and ridge/valley proxies**

To capture roof topology beyond plane counts and slopes, plane adjacency is derived from shared edges between faces. An edge map is created over roof faces; each undirected edge is stored using a canonical orientation (e.g., smaller vertex index first) to make adjacency deterministic. Two planes are adjacent if any face from plane A shares an edge with any face from plane B. Shared-edge lengths are accumulated per plane pair and aggregated per roof.

For each shared edge between two planes, a dihedral sign proxy is computed using plane normals and a consistently oriented edge vector. Edges classified as convex act as ridge proxies and concave edges act as valley proxies. These proxies support explainable complexity assessment and can be used for detailing-related quantity rules (e.g., ridge/valley flashing length), even if the present paper focuses primarily on area-driven roof layers and structural timber.

**Roof metrics extraction**

For each building roof, we compute a set of metrics used both for classification and for quantity estimation:

- Roof surface area: sum of roof plane areas.
- Horizontal projection area: projection of roof surfaces onto the horizontal plane; implemented to support polygon holes (outer ring minus inner rings) where applicable.
- Slope metrics: maximum slope of the steepest plane (and optional area-weighted summaries).
- Plane count: number of roof planes after merging.
- Span proxy: derived from the building footprint, obtained by selecting the largest near-base horizontal ring and computing a conservative span proxy (typically the shorter side of the footprint bounding rectangle).

Quality indicators (QA) are computed to support interpretation (especially for LoD3): non-manifold edge counts and tiny-plane statistics (number and area ratio of very small planes). These indicators do not change the LoD2 operational method, but help explain LoD3 fragmentation effects.

**Rule-based roof type classification**

Roofs are classified into a construction-oriented typology set: gable, shed, complex, flat, and hip (Figure 1). The classifier is intentionally rule-based to remain interpretable and auditable at city scale and to support downstream normative mapping and material computations.


*Figure 1: Roof types for automatic classification.*

The concept used for the decision logic is as follows:

1. Detect flat roofs based on near-horizontal slope conditions (including robust handling of small steep facets);
2. Detect shed roofs by a single dominant pitch/orientation;
3. Detect mansard roofs by the coexistence of shallow and steep planes above area thresholds;
4. Separate gable and hip using plane orientation dominance and adjacency structure; and
5. Assign complex as a conservative fallback when rules are inconclusive or roofs are irregular/fragmented.

LoD2 is used as the primary input because its roof geometry is typically less fragmented and more stable for plane merging and rule triggers. LoD3 introduces additional small facets (dormers, chimneys, parapets), inflating plane counts and disturbing dominance logic, therefore LoD3 is treated as comparative evidence.

**Mapping roof types to normative structural classes**

To estimate timber quantities, each roof is mapped to a normative structural class based on roof type and the span proxy L. The mapping follows normative span intervals and a deterministic tie-breaking strategy: where multiple classes cover the same span, the class with minimum timber consumption is selected; gaps are resolved by selecting the next stronger admissible class; spans above the maximum use the strongest available class as an upper approximation.

For hip roofs, timber consumption is approximated using gable-roof normative rules at this level of abstraction. Complex roofs are treated as gable-equivalent for baseline normative selection by span, with an additional correction factor representing geometric complexity (derived from concave-edge intensity proxies).

**Roof structure's material quantity estimation**

Timber volume is computed as:

$$V_{timber} = A_{proj} \times r_{norm} \times k \quad (1)$$

where $A_{proj}$ is the horizontal projection of the roof's area, $r_{norm}$ is the normative timber consumption rate (m³ per m² of horizontal projection) for the selected class, and $k$ is a correction factor (k = 1 by default; k > 1 for complex roofs based on complexity bins).

This formulation couples geometry-derived area and span to normative structural selection and complexity-aware adjustments.

Concrete quantity in flat roof structures is computed much more straightforward, as:

$$V_{concrete} = A_{roof} \times d_{slab} \quad (2)$$

where $A_{roof}$ is the roof's area, and $d_{slab}$ is the concrete slab thickness.

Pseudocode below provides the step-by-step operational detail.

*FOR each Building in CityJSON:*

*1) Extract roof faces*

- *Prefer semantic RoofSurface faces*
- *Else fallback: upward-facing faces near top elevation*

*2) Compute per-face geometry*

- *normal, area, centroid, horizontal projection area (holes supported)*

*3) Merge faces into roof planes*

- *cluster by (normal angle tolerance) AND (coplanarity distance tolerance)*
- *compute plane area + plane slope*

*4) Build plane adjacency + ridge/valley proxies*

- *construct deterministic shared-edge map (canonical edge key)*
- *for each shared edge between different planes:*
  - *accumulate shared-edge length*
  - *classify as convex (ridge proxy) or concave (valley proxy)*

*5) Compute roof metrics + QA*

- *roof area, projected area, max slope, plane count*
- *footprint span proxy from largest near-base horizontal ring*
- *QA: non-manifold edges, tiny-plane statistics*

*6) Classify roof type (rule-based)*

- *decide flat vs non-flat from slope*
- *detect shed / gable / hip using plane counts, dominant plane areas,*
- *slope distribution, concave ratio, and footprint aspect ratio*
- *else fallback: complex*

*7) Export per-building results (roof_type + metrics)*

**Implementation notes and reproducibility**

All processing is deterministic with explicit parameter thresholds (e.g., plane normal tolerance, coplanarity tolerance, minimum shared-edge length, and slope cutoffs). Intermediate outputs (roof planes, QA indicators, roof metrics) are exported alongside final labels and quantities to support diagnosis and sensitivity analyses, enabling reproducibility and auditability.

## Data and case study

**Study area and datasets**

The case study is conducted on a dense, mixed building block in Belgrade (Serbia), selected as a representative urban fabric with a wide variety of roof geometries (flat roofs, mono-pitch roofs, and multi-patch roofs with local superstructures). The study area is used to demonstrate an end-to-end workflow for:

- Automatic roof type inference from an open 3D city model,
- Roof metrics extraction, and
- Rule-based material quantity assessment for roof layers and the timber roof structure.

Two CityJSON datasets are used for the same spatial extent: a LoD2 model, where roof geometry is represented with a cleaner set of roof facets, and a LoD3 model, which additionally encodes façade openings and finer roof superstructures (Figure 2). While the classifier can be executed on both datasets, LoD2 is treated as the primary operational input, and LoD3 is used to illustrate how increased geometric detail can affect planar fragmentation and typology stability.

The evaluation dataset contains 88 buildings with manually assigned roof type labels. A subset of six buildings was further modelled in BIM to obtain baseline take-offs for validating roof structure quantities and timber volumes (Figure 3).

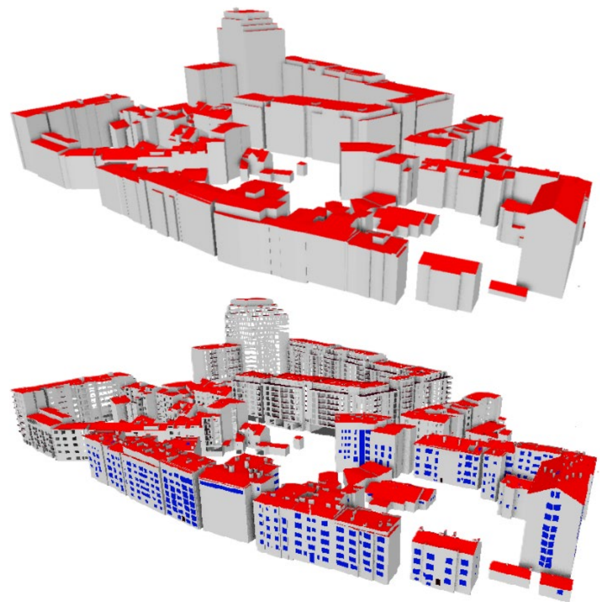

*Figure 2: 3D City models of the case area at various level of detail: CityJSON models at LoD2 (top), and LoD3 (bottom).*

**3D city models (CityJSON LoD2 and LoD3)**

Both city models follow the CityJSON data model and represent buildings as boundary representations with triangulated surfaces. Roof faces are semantically identifiable (RoofSurface) and are used as the primary input for plane merging, metric extraction, and roof type classification. The two levels of detail differ mainly in the degree of geometric fragmentation: LoD2 tends to express roofs as a smaller number of planar patches, whereas LoD3 introduces additional small facets (e.g., dormers, chimneys, parapets), which increases the number of planes after merging and may perturb dominance-based rules. This LoD effect is reported in the Results section through LoD2–LoD3 prediction agreement–

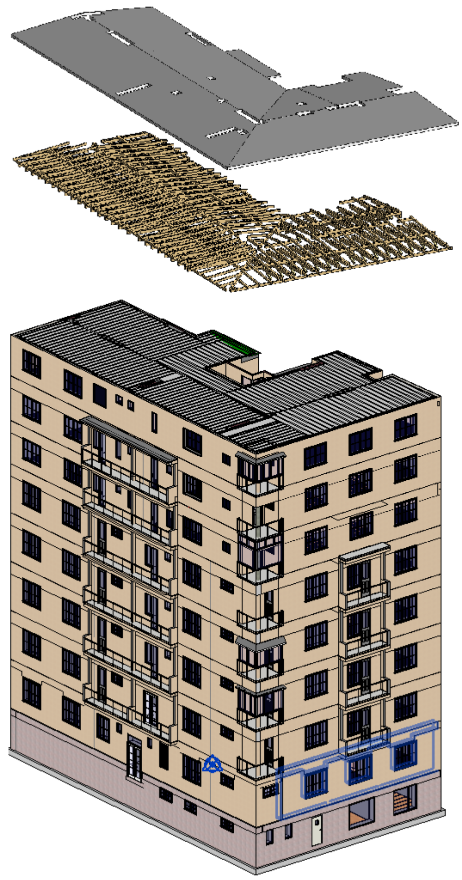

*Figure 3: BIM model of a residential building from the case area needed for validation of a roof structure's material quantity computation.*

## Results

**Roof type classification**

The evaluation uses 88 labelled buildings from the Belgrade test area. Using CityJSON LoD2 as input, the classifier achieves an overall accuracy of 0.716 and a macro-averaged F1-score of 0.543. Running the same code on the LoD3 model yields an accuracy of 0.557 and macro-F1 of 0.403. Agreement between LoD2 and LoD3-based predictions is 0.534 (Cohen's $\kappa = 0.375$), indicating that the choice of LoD representation affects typology stability even when the algorithm is unchanged.

Table 1 (LoD2) and Table 2 (LoD3) report per-class precision/recall/F1 against manual labels, while Table 3 and Table 4 provide the corresponding confusion matrices. Across both LoDs, errors concentrate on (i) flat vs non-flat ambiguity and (ii) gable vs complex ambiguity.

**Extracted roof metrics and their role in the pipeline**

For each building, the workflow reports plane count, maximum roof slope, roof area, horizontal projection area, and a footprint span proxy. These metrics provide the inputs to the material quantity rules and enable explainable links between geometry, typology, and quantities.

*Table 1: Roof type classification performance using CityJSON LoD2 as input (automatic predictions vs manual labels). Support denotes the number of labelled buildings per class; macro-averages are computed across classes.*

| Roof type | Support (n) | Precision | Recall | F1 |
|---|---|---|---|---|
| flat | 16 | 1.000 | 0.562 | 0.720 |
| shed | 24 | 1.000 | 0.833 | 0.909 |
| gable | 12 | 0.400 | 0.333 | 0.364 |
| hip | 1 | 0.000 | 0.000 | 0.000 |
| complex | 35 | 0.625 | 0.857 | 0.723 |
| Macro avg | 88 | 0.605 | 0.517 | 0.543 |
| Overall accuracy | 88 | | | 0.716 |

*Table 2: Table 2. Roof type classification performance using CityJSON LoD3 as input (same code and rules; automatic predictions vs manual labels). Differences relative to Table 1 quantify sensitivity to LoD representation.*

| Roof type | Support (n) | Precision | Recall | F1 |
|---|---|---|---|---|
| flat | 16 | 0.643 | 0.562 | 0.600 |
| shed | 24 | 0.647 | 0.917 | 0.759 |
| gable | 12 | 0.000 | 0.000 | 0.000 |
| hip | 1 | 0.000 | 0.000 | 0.000 |
| complex | 35 | 0.737 | 0.400 | 0.519 |
| Macro avg | 88 | 0.405 | 0.376 | 0.403 |
| Overall accuracy | 88 | | | 0.557 |

*Table 3: Confusion matrix for LoD2-based classification (rows: manual labels; columns: automatic predictions).*

| Pred flat | Pred shed | Pred gable | Pred hip | Pred complex |
|---|---|---|---|---|
| 9 | 0 | 0 | 0 | 7 |
| 0 | 20 | 2 | 0 | 2 |
| 0 | 0 | 4 | 0 | 8 |
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 4 | 1 | 30 |

*Table 4: Confusion matrix for LoD3-based classification (rows: manual labels; columns: automatic predictions).*

| Pred flat | Pred shed | Pred gable | Pred hip | Pred complex |
|---|---|---|---|---|
| 9 | 5 | 0 | 0 | 2 |
| 0 | 22 | 0 | 0 | 2 |
| 0 | 9 | 0 | 0 | 3 |
| 0 | 1 | 0 | 0 | 0 |
| 5 | 6 | 0 | 0 | 24 |

**Implications for material quantities**

Using identical quantity rules, the mean absolute difference between automatic and manual estimates is 4.32 m³ for timber (mean relative error 0.033) and 7.80 m³ for concrete (mean relative error 0.385, computed where manual concrete > 0). Table 6 summarizes this sensitivity.

Concrete quantities are particularly sensitive to the flat/non-flat decision: 5 of 14 manually labelled flat roofs are classified as non-flat in the LoD2-based predictions and 8 of 14 in the LoD3-based predictions (Table 5). Because concrete quantities are only computed for flat-roof structural archetypes, these misclassifications force the concrete estimate to zero and therefore dominate concrete errors.

For the BIM validation subset (n=6), Table 7 compares automatic and manual rule-based quantities against BIM-derived baselines. Across buildings where a BIM baseline is available, the mean relative error of the automatic estimate is 0.647 for timber and 0.600 for concrete (where BIM baseline > 0). Remaining discrepancies are expected to reflect both typology errors and differences in modelling scope between rule-based archetypes and BIM take-offs.

*Table 5: Agreement between LoD2- and LoD3-based automatic roof type predictions, including flat-roof sensitivity. Cohen's κ complements exact agreement by adjusting for chance agreement.*

| Metric | Value |
|---|---|
| Exact agreement (LoD2 vs LoD3 predictions) | 0.511 |
| Cohen's κ (LoD2 vs LoD3 predictions) | 0.335 |
| Flat roofs misclassified (LoD2 predictions) | 1/10 |
| Flat roofs misclassified (LoD3 predictions) | 3/10 |

*Table 6: Sensitivity of rule-based material quantities to roof type classification (automatic vs manual labels, same quantity rules). Concrete quantities are evaluated only for roofs mapped to flat-roof structural archetypes.*

| Material / group | MAE (m³) | Mean rel. error | Interpretation |
|---|---|---|---|
| Timber (all buildings) | 3.22 | 0.023 | Stable for pitched roofs; differences mainly driven by typology/normative-type mapping |
| Concrete (flat-roof archetypes only) | 15.48 | 0.400 | Highly sensitive to flat/non-flat decision; misclassification forces concrete estimate to 0 |

*Table 7: BIM validation subset (n=6): automatic and manual rule-based quantities compared against BIM-derived baseline take-offs. Remaining differences may reflect both typology errors and modelling-scope differences (e.g., inclusion of secondary timber members).*

| Building ID | Manual roof type | Auto roof type (LoD2 pipeline) | # planes | Max slope (°) | Timber Auto ($m^3$) | Timber Manual ($m^3$) | Timber BIM ($m^3$) | Concrete Auto ($m^3$) | Concrete Manual ($m^3$) | Concrete BIM ($m^3$) |
|---|---|---|---|---|---|---|---|---|---|---|
| Cv_90 | complex | complex | 5.0 | 6.54 | 22.87 | 22.87 | 19.28 | 0.0 | 0.0 | 0.0 |
| Dl_55 | gable | complex | 6.0 | 32.85 | 15.64 | 17.37 | 16.48 | 0.0 | 0.0 | 0.0 |
| StGl_34 | gable | complex | 5.0 | 29.83 | 13.17 | 10.98 | 4.07 | 0.0 | 0.0 | 0.0 |
| StGl_36 | complex | complex | 9.0 | 45.91 | 16.47 | 16.47 | 18.57 | 0.0 | 0.0 | 0.0 |
| Dl_28 | flat | complex | 6.0 | 18.16 | 36.23 | 0.0 | 0.0 | 0.0 | 68.11 | 67.14 |
| Dl_57 | flat | flat | 2.0 | 2.54 | 0.0 | 0.0 | 0.0 | 8.42 | 8.42 | 10.52 |

# Discussion

### Interpreting classification performance and error structure

The results suggest that an interpretable, rule-based classifier can support operational roof typology inference at city scale when applied to a semantically structured LoD2 CityJSON model. On the 88-building dataset, LoD2 classification achieves 0.716 overall accuracy with a macro-averaged F1 of 0.543, indicating that several roof types exhibit distinctive planar signatures, while performance remains uneven across classes.

Two ambiguity families dominate. Flat vs non-flat errors arise when small steep facets on otherwise flat roofs (e.g., parapets or technical housings) affect slope-based triggers even after robustness filtering. Gable vs complex errors occur for multi-part roofs (annexes, terraces, multiple ridges), where “complex” acts as a conservative fallback suggesting that a single roof archetype may be insufficient without segmentation.

### Why LoD2 is operationally preferable, and what LoD3 adds

LoD2 is operationally preferable because it preserves the primary roof shape with less fragmentation, while LoD3 adds superstructures and small facets that perturb dominance-based rules. This representational sensitivity is visible in the LoD2–LoD3 agreement statistics (0.511 exact agreement, $\kappa = 0.335$), indicating only moderate stability when the algorithm is unchanged but the geometric representation differs. LoD3 is therefore best treated as diagnostic evidence motivating generalization/feature suppression or a “structural roof surface” abstraction.

### Effects of typology errors on material quantities

Typology errors propagate asymmetrically into quantities. Timber estimates are comparatively stable because most non-flat types map to pitched-roof families and are driven primarily by projection area and span-based normative selection; this is reflected in a timber MAE of 3.22 $m^3$ (mean relative error 0.023) between automatic- and manual-driven estimates under identical quantity rules.

Concrete is highly sensitive because it is computed only for flat-roof archetypes; misclassifying a flat roof as non-flat forces the concrete estimate to zero and creates step-change errors. Accordingly, concrete shows a larger MAE of 15.48 $m^3$ (mean relative error 0.400, computed where manual concrete > 0). For city-scale accounting, improving the flat/non-flat decision (e.g., area-weighted flatness indicators rather than max-slope triggers) is likely to yield larger downstream gains than refining pitched-roof subtypes.

### Interpretable rules vs. end-to-end deep learning

The rule-based approach is appropriate here because it provides traceability from geometry → typology → archetype → quantities, is data-efficient when labelled city-scale roof datasets are limited, and aligns naturally with normative quantity rules. Deep learning remains a useful complement, e.g., as a lightweight model applied only to ambiguous cases (near flat thresholds or “complex” candidates) or as a calibration layer for threshold tuning.

### Limitations and targeted improvements

Three limitations follow from the error modes: (1) single label per building roof, which motivates roof-instance segmentation; (2) fragmentation sensitivity, suggesting area-weighted flatness indicators rather than max-slope triggers; and (3) validation scope, since the BIM subset (n=6) supports feasibility but should be expanded for stronger generalisation claims.

# Conclusions

This paper presented an interpretable CityJSON-based workflow for automated roof type classification and roof metrics extraction, enabling rule-based estimation of roof material quantities at city scale. Using a Belgrade test area, the approach demonstrated that a rule-driven classifier can provide reliable roof typology inference when applied to a semantically structured LoD2 city model, while simultaneously producing the geometry-derived parameters required for downstream quantity rules (roof areas, projected areas, slope statistics, plane counts, and a footprint-based span proxy).

The results show that classification errors are structured rather than random. Misclassifications concentrate on flat versus non-flat ambiguity caused by small steep facets on otherwise flat roofs and gable versus complex ambiguity for irregular or multi-part roofs. Importantly, the study quantified how such typology errors propagate to quantity estimation: timber volume estimates remain comparatively stable across non-flat typologies, whereas concrete quantities are highly sensitive to flat roof recall because flat/non-flat decisions control whether the flat-roof archetype branch is applied at all.

A key methodological takeaway is that LoD selection matters. While the same classifier can be executed on LoD3, LoD2 provides a more stable basis for city-scale roof typology inference due to reduced geometric fragmentation. LoD3 is therefore best interpreted as a complementary dataset for diagnosing representation-induced instability and motivating generalization and segmentation strategies.

## Future work

Future work will prioritize:

1. Roof instance segmentation to handle buildings with multiple roof components (main roof vs annex), reducing “complex” fallbacks and improving quantity mapping;
2. Lod3 generalization to suppress/aggregate superstructure facets into a stable “structural roof surface” prior to classification; and
3. Improved flat-roof robustness using area-weighted flatness indicators to reduce flat/non-flat errors that dominate concrete quantities.
4. In addition, we will expand validation beyond six BIM baselines and report uncertainty via sensitivity analysis over key thresholds and normative mappings.

## Acknowledgments

This research was funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101082068 (CircBoost project).

## References

Biljecki, F., & Dehbi, Y. (2019, September). Raise the roof: towards generating lod2 models without aerial surveys using machine learning. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-4/W8, 27–34. doi:10.5194/isprs-annals-IV-4-W8-27-2019

Buyukdemircioglu, M., Can, R., & Kocaman, S. (2021, June). Deep learning based roof type classification using very high resolution aerial imagery. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B3-2021, 55–60. doi:10.5194/isprs-archives-XLIII-B3-2021-55-2021

Dušan Isailović, D. I., Ana Nadaždi, A. N., Aleksandra Parezanović, A. P., Zorana Petojević, Z. P., & Nenad Višnjevac, N. V. (2025, July). 3d city model conversion to bim for automatic material quantities assessment. doi:10.35490/EC3.2025.378

Janisio-Pawłowska, D., & Pawłowski, W. (2024, January). Implementation of bim data in citygml—research and perspectives for creating a qgis plugin for spatial analysis: experience from poland. Sustainability, 16, 642. doi:10.3390/su16020642

Ledoux, H., Arroyo Ohori, K., Kumar, K., Dukai, B., Labetski, A., & Vitalis, S. (2019, December). CityJSON: a compact and easy-to-use encoding of the CityGML data model. Open Geospatial Data, Software and Standards, 4, 4. doi:10.1186/s40965-019-0064-0

Mohajeri, N., Assouline, D., Guiboud, B., Bill, A., Gudmundsson, A., & Scartezzini, J.-L. (2018, June). A city-scale roof shape classification using machine learning for solar energy applications. Renewable Energy, 121, 81–93. doi:10.1016/j.renene.2017.12.096

Open Geospatial Consortium. (2026, April 7). OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard, Version 3.0, OGC Doc No. 20-010, 2021. Retrieved from OGC: https://docs.ogc.org/is/20-010/20-010.html

Parezanović, A., Nadaždi, A., Isailović, D., Višnjevac, N., & Petojević, Z. (2025, April). Mapping the urban building stock for a circular economy by integrating GIS and BIM. A case study from Belgrade, Serbia. Resources, Conservation and Recycling, 215, 108075. doi:10.1016/j.resconrec.2024.108075

Pei, W. Y., Biljecki, F., & Stouffs, R. (2022, October). Dataset for urban scale building stock modelling: identification and review of potential data collection approaches. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4/W2-2022, 225–232. doi:10.5194/isprs-annals-X-4-W2-2022-225-2022

Sitong, L., Chengzhi, R., Chi, Z., & Qiuxian, W. (2023). Building and roof type classification on 3D city models. Building and roof type classification on 3D city models. Retrieved February 8, 2026, from https://repository.tudelft.nl/record/uuid:585100a5-f696-487f-a197-858ffa422f26