2023 European Conference on Computing in Construction
40th International CIB W78 Conference
Heraklion, Crete, Greece
July 10-12, 2023

PREDICTING RECOVERABLE MATERIAL STOCK IN BUILDINGS: USING MACHINE LEARNING WITH PRE-DEMOLITION AUDIT DATA AS A CASE STUDY

Natalia Ewa Kobylinska*, Deepika Raghu, Matthew Gordon, Jens Hunhevicz,
Catherine De Wolf
ETH Zürich, Switzerland
*corresponding author: kobylinska@ibi.baug.ethz.ch

Abstract

Early specification of materials in buildings before their demolition could foster reuse in the construction industry. Studies have already shown the usefulness of machine learning in demolition waste estimation; however, application to real-world datasets is still limited. This study tests the feasibility of predicting recoverable material stock in the local context of the city of Zurich. The results show promise for the overall approach, although training models by using a small and heterogeneous dataset poses challenges. Therefore, we conceptualized an improved demolition data collection, processing, and dissemination. The resulting framework could help researchers and authorities in urban material stock estimation.

Introduction

The building floor area is expected to rapidly expand in the next couple of decades, even in already densely urbanized parts of the world like Europe (UN Environment and International Energy Agency, 2017). This trend raises questions about the continuous generation of construction and demolition waste and the growing demand for raw materials in new buildings. Indeed, the construction industry is already one of the most significant carbon emitters and waste producers globally (Akhtar and Sarmah, 2018; European Commission, 2016). Using the retiring building stock as a mine for secondary materials for new construction would help the industry lower its environmental impact. To do so would require information about the suitability for recycling and reusing materials in existent buildings before their decommission. Such information could help organize a circular project's logistics, estimate demolition and recycling costs, and prioritize interventions of reuse agents.

Unfortunately, existing buildings are rarely represented, e.g. in BIM or CAD drawings. Reconstructing an existing building's inventory, either manually or with the help of advanced technological methods like scan-to-BIM, tends to be time- and resource-intensive. Such reconstruction often has severe limitations, for example when elements are hidden under a building's outer layer (Honic et al., 2021). A more scalable approach to characterize urban material stocks is the bottom-up development of so-called archetypes, often categorized by e.g. a building's age and its primary function (e.g. TABULA WebTool, 2015). The developed archetypes are sometimes extrapolated to an urban scale to estimate a city's material stock and predict material flow (Heeren and Hellweg, 2018; Ostermeyer *et al.*, 2018). Their development requires thorough in situ visits and access to detailed building documentation. The final estimation relies on the granularity of the developed typologies.

As an emerging approach, we identified three studies that explore data-driven modeling for detecting material presence and estimating bulk waste in buildings. Akanbi *et al.* (2020) developed a deep learning algorithm that predicts the amount of demolition waste in three categories: reusable, recyclable, and disposable. The trained model exhibits a strong skill, but it requires information on a building's structural material as an input. This information is not always straightforward to obtain. Moreover, the model's output lacks sufficient granularity for planning material recovery (i.e. for amounts of specific materials). Cha *et al.* (2020) presented a methodology using a random forest machine-learning algorithm to estimate demolition waste per material type, usable for small datasets and with mixed (i.e. continuous and categorical) inputs. Both studies trained models on labeled datasets, obtained either from the private sector or during previous research efforts. In contrast, Wu *et al.* (2022) propose public datasets (Gothenburg and the Stockholm City Archives) as a possible source of building material information to create a dataset to predict hazardous materials. Nevertheless, their model output is only binary (i.e. a hazardous material type is detected or not detected) and does not predict the material stock composition.

This study builds on the above research to develop a data-driven material stock estimation, applicable to available datasets in the local context of Zurich. We merge open-

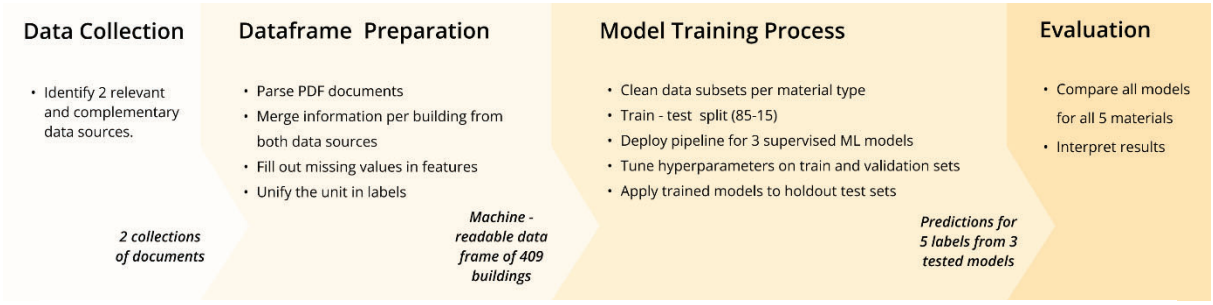

Figure 1: The four methodological steps.

access cadastre data with local semi-open demolition audit records into a new dataset of 409 residential buildings. We then trained three types of algorithms: linear regression (LR), random forest regressor (RFR), and the extreme gradient boosting (XGBoost). The introduced proof of concept allows for the prediction of amounts of wood, mineral, metal, glass, and roof tile materials in the residential building stock. The available data quality sometimes posed challenges, as was reflected in the results. Therefore, the paper conceptualizes a framework for data collection, processing, and dissemination to help establish a more structured, quality-assured, and up-to-date material stock dataset at the urban level.

Methodology

The scope of this study is limited to residential buildings raised between 1850 and 1973 in Zurich, a time frame based on data available. We used a residential sample because, according to Zurich's Statistics Office (Stadt Zürich, 2022), more residential than non-residential buildings were demolished in the city in recent years. Furthermore, the ability of machine learning (ML) models for generalization (i.e. to adapt prediction to new instances) was assumed more challenging on a non-residential sample because commercial, educational, or office buildings usually exhibit a higher variety of spatial designs.

Four methodological steps helped estimate recoverable amounts of materials per building (visualized in Fig.1): (1) data collection, (2) data frame preparation, (3) the model training process, and (4) evaluation of the models. The first three steps are discussed below; step 4 is discussed in the results section.

Data Collection

In supervised machine learning, a labeled dataset is used to train a model. Features (X) are independent variables in the dataset, used as input to make a prediction. Labels (Y) are dependent variables – values that one wants to predict. In this study, a custom collection of relevant building attributes serves as features X, and amounts of specific materials (in metric tons) become labels Y. To create a machine-readable dataset, consistent records on both X and Y needed to be identified and merged. Ideally, features X needed to contain openly accessible information to ensure that a trained model can be easily used for an instant estimation of the amount of materials in a building. Characterization of two data sources used in this study is provided in Table 1.

The *Gebäude- und Wohnungsregister* (GWR, or the Federal Register of Buildings and Dwellings) grants public access to a set of attributes for all buildings in Switzerland. The attributes can be queried by the *Eidgenössische Gebäudeidentifikator* (EGID, or the Federal Building Identification Number) which can be found by a building's postal address. The features X of a building were extracted from this database with a custom Python script and included: footprint area, gross volume, year of construction, period of construction, number of stories, and number of apartments.

In parallel, the *Umwelt- und Gesundheitsschutz Zurich* (UGZ, or the Office of Environmental and Health Protection Zurich) provided access to disposal concepts as a source of information for labels Y. Disposal concepts (as .pdf or .jpg formats) must be submitted to the UGZ office before a building is renovated or demolished. The focus of this process is to specify expected hazardous materials, but documents also need to include a table with types, volumes, and/or weights of non-polluted materials in a building. The estimation is made by an expert conducting a building audit. This study used all available records on full demolitions of residential buildings between 2018 and 2022. The final dataset acquired from UGZ equaled 206 demolition projects.

Data Preparation

Among the 206 demolition projects acquired from UGZ, there were multiple cases of more than one building per project (such as demolition of a whole neighborhood).

Table 1: Characterization of the two used data sources.

Source	UGZ	GWR
Relevant information available	Weight and/or volume for 2 to 40 material types; images of buildings; address	Footprint area, gross volume, year and period of construction, stories, and apartments count
Data format and resolution	.pdf or .jpg; data per demolition project	.csv; data per building
Set size	206 projects	> 400,000 buildings

Additionally, only 124 instances followed a material table format recommended for disposal concepts, see Entsorgungskonzept Rück- und Umbau, 2020. Even among these instances, most customized the format, thus impeding the use of automated extraction of information. Such inconsistency in data format required a manual and time-consuming data parsing process.

The quality of records also strongly varied. The number of materials declared in different demolition projects ranged from 2 to 40 different types. The presence of some materials was not recorded enough times in the whole sample for it to be useful in the ML training (e.g., gypsum). Other materials (e.g. contaminated wood or road rubble), were of no interest for this study's focus on recovering building materials. Some categories were defined too broadly (e.g. as unsorted mixed waste or burnable waste). Finally, five labels were chosen for Y: *wood, metal, roof tile, glass* and *mineral* ('mineral' defined as a mixture of exclusively mineral waste such as concrete, brick, sand lime, and natural stone; see BAFU, 2006). Not all 206 projects had information on all five chosen labels, which resulted in different sizes of training sets per material.

To arrive at a structured data frame with the labels Y in a uniform unit, the inconsistent format of records required making assumptions (listed in Table 2). To complete the datasets, a roof type (flat, mansard, pitched, or mixed) was manually assigned to a building based on photos in UGZ records. The hypothesis was that different roof types could significantly affect the amount of waste (especially *wood*) obtained from a building.

Finally, missing volume attributes were filled by importing publicly available CityGML LoD 2.3 models of Zurich (geocat.ch, 2018) into the Rhino environment and calculating the volume with a custom Grasshopper script. The final data frame constituted 409 data points with the characteristics presented in Table 3.

Table 2: Necessary assumptions made to prepare a consistent data frame.

Inconsistency	Assumption
Volume (V [m^3]) was sometimes recorded in 'compact' and sometimes in 'loose' categories	V_loose = 1.3 ∗ V_solid
Some demolition projects encompassed auxiliary buildings (i.e. sheds, garages)	V_ residential = 0.9 ∗ V_total
Amounts recorded as volume needed to be converted to mass	The assumed density of a material = average density of corresponding materials (see KBOB, 2016)
A demolition project encompassed more than one residential building and Y values were aggregated per project	Y_individual_building = Y_demolition_project * the building's percent volume contribution

Model Training

Predicting a continuous value Y from the set of features X is a regression problem. For this study, we performed a regression on a heterogeneous and small dataset that was limited by data availability. These boundaries determined the choice of the ML models and the overall training strategy. Two tree ensemble models, namely Random Forest Regressor (RFR) and XGBoost (XGB), were tested and compared with a third model - linear regression (LR). All three models were trained in Google Colab notebook using Python and Scikit-learn library tools. The choice of RFR was considered applicable to the problem for two reasons: it can handle mixed input (categorical and continuous data) and it is also applicable to a small sample size (Cha et al., 2020). The second ensemble tree algorithm used in this study (XGB) uses boosting instead of bagging technique while combining results from N learners into the final result. The LR served as a baseline model for the performance evaluation of the other models.

Only 140 of the total 409 buildings had information on all five investigated materials at once, meaning that one or more labels per building were missing for most cases. Reducing the training set to 140 instances was expected to extremely compromise the models' performance. Instead of developing a model which would predict all the labels at once, separate models for every single label were developed.

To clean the data frame for ML training, we first eliminated 'not a number' (NaN) values in the dataset's labels. Then, we handled outliers per each continuous feature with the interquartile range (IQR) method. IQR equals a difference between the third and the first quartile of a sample (IQR = Q3 – Q1).The values bigger than Q3 + 1.5 ∗ IQR or smaller than Q1 – 1.5 ∗ IQR were consequently dropped from the set. The final number of data points used for training the models is summarized in Table 4.

Table 3: Characterization of features and labels in the assembled data frame of 409 buildings.

Features (X)/Labels (Y)	Data type	Unit/Count
X: Gross volume	continuous	m^3
X: Footprint area	continuous	m^2
X: Apartments count	continuous	-
X: Stories count	continuous	-
X: Location (district)	categorical	12 categories
X: Location (zipcode)	categorical	21 categories
X: Year of construction	continuous	-
X: Period of construction	categorical	6 categories
X: Roof type	categorical	4 categories
Y: wood/ metal/ roof tile/ glass/ mineral	continuous	t

Table 4: Number of data points used for training supervised machine models, per material type.

wood	metal	roof tile	glass	mineral
286	277	213	283	309

Next, exploratory data analysis was conducted to reveal important statistical characteristics of a sample and correlations between features (the results of the pre-training sample analysis are described in the results section). The Pandas library and the One-Hot Encoding method were used to encode categorical features as machine-readable binary vectors. Additionally, feature scaling was implemented in the LR model due to its sensitivity to non-normalized inputs. In all the models, splitting the dataset into test and train sets posed a challenge due to the size and heterogeneity of the data. Even though the data is always shuffled before splitting, in small datasets the results on a test set can be biased if the test set has different distribution from a train set. To address this limitation, the Kolmogorov-Smirnov (K-S) statistical test was implemented. It allowed us to pick a split that ensured relative statistical similarity between the test and train sets for all the models. The holdout test set always constituted 15% of the sample (regardless of material type) and was used as the final estimation of models' performance, after their training and validation.

In two ensemble tree models, a hyper-parameters search was performed in a Stratified K-Fold Cross-Validation on the remaining 85% of the sample. Cross-validation was used with ten folds for *metal*, *mineral waste*, and *wood*, and with five folds for the two smaller samples (*glass* and *roof tile*). The baseline LR model did not require hyperparameters tuning and therefore does not have a separate validation set. A tree ensemble model training steps can be followed in Fig. 2, on the example of the *wood* sample.

Model Evaluation

In the final step, models were evaluated on the holdout test set by two chosen metrics: R-squared and mean absolute error (MAE). R-squared metric needs to be maximized and MAE needs to be minimized (y_i = true value, $\overline{y}$ = mean true value, y_i-hat = predicted value, n = sample size), as displayed in Eq. (1) and (2).

$$R^2 = 1 - \frac{(y_i - \hat{y}_i)^2}{(y_i - \overline{y})^2} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n} \quad (2)$$

Results

Exploratory Data Analysis

The results showed that most of residential buildings demolished in Zurich from 2018 to 2022 had an average life cycle of 70 to 90 years. Most of the sample represents single-family houses (1 apartment) and multi-family houses (6–7 apartments). A typical building's volume is 1500 to 2000 m^3 and a typical footprint is 170 m^2. Before outlier removal, the distribution of all continuous features was strongly right-skewed (i.e., most of the sample was in a low-value area with strong outliers in a high-value area). It was also observed how imbalanced the classes in categorical features were. Specific locations in the city (Kreis 8 or 12 and zip codes 8003, 8008, 8045, and 8051) were highly underrepresented due to the specific characteristic of the collected sample available at UGZ.

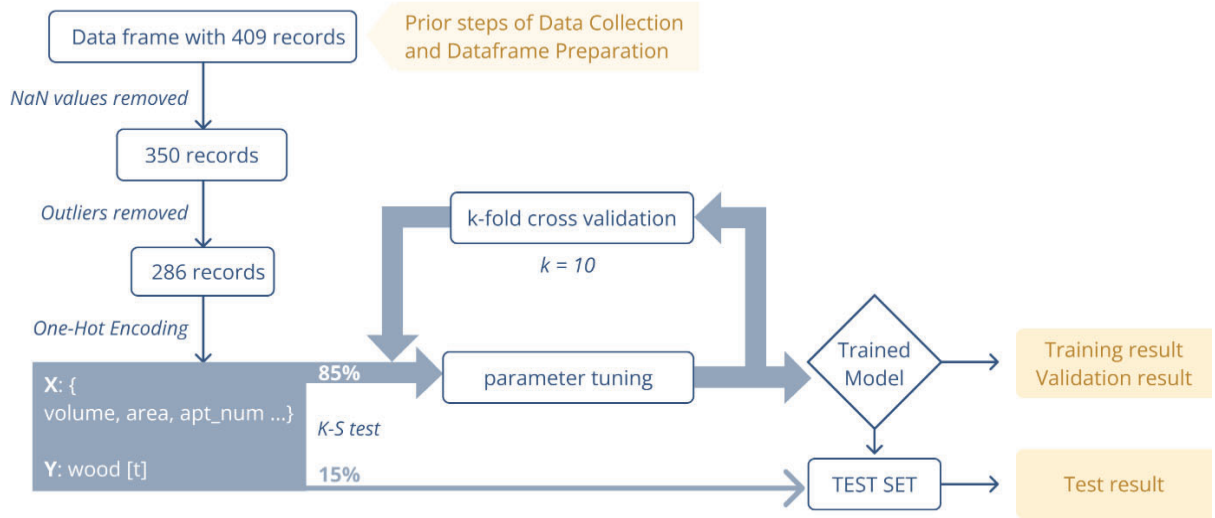

Figure 2: A machine-learning tree ensemble model training process visualized on the example of the wood sample.

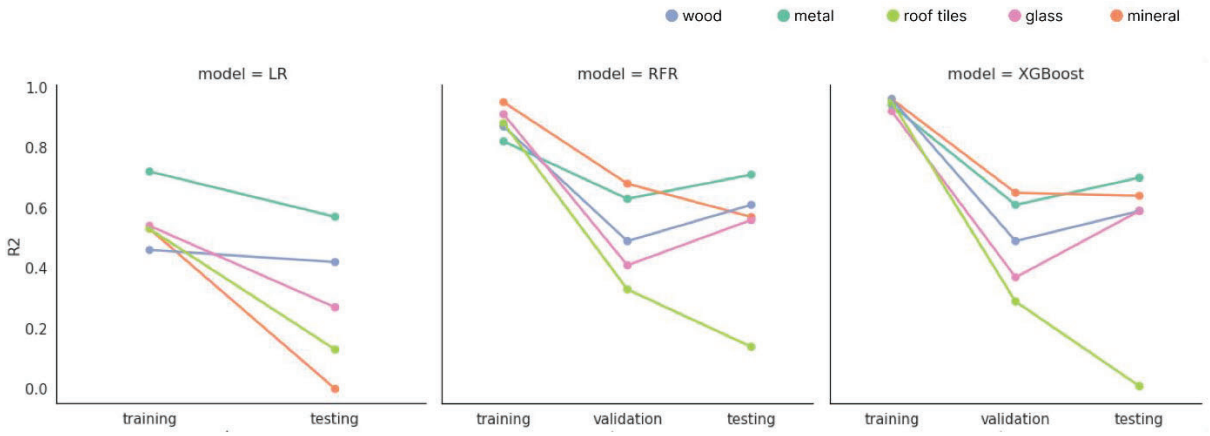

Figure 3: Comparison of the models' performance across the five investigated materials using the R-squared metric. Results are shown for the training, validation, and testing stage.

The strongest linear correlation for all the materials was the one with the volume feature. *Wood* and *mineral* were also correlated with the footprint area. *Mineral* amounts seemed to be bigger, the newer the buildings were. Linear correlation between other materials and buildings' year of construction or a roof type was much weaker or non-existent.

Material Prediction

The overall results across the material datasets and trained models are evaluated with R-squared and MAE metrics (represented in Fig. 3 and Table 5).

Next to the achieved performance on the test sets, the context of training and validation results helps to assess a model's skill. A significant difference between training and testing results is usually indicative of the model's overfitting (high variance error). On the contrary, a small difference, but poor performance on both sets, is usually indicative of underfitting (high bias error). Both tree ensemble models outperform the LR model, but exhibit overfitting of the data, which is especially pronounced in the case of the XGBoost model.

Prediction of *mineral, wood, metal,* and *glass* quantities predicted by XGB overall proved the most successful, with R-square between 0.59 and 0.70. However, the results from the RFR model follow those of XGBoost very closely with the better generalization pattern. The assessment with MAE metric speaks slightly in favor of the RFR model, however the difference in the results from both tree ensemble algorithms is small.

Table 5: Mean absolute error (MAE) for predictions from all models, for a test set only, across all the materials. All values are expressed in metric tons and can be interpreted per sample, relative to its standard deviation (std, in grey). The best result per sample is marked in green.

MAE [t]	wood	mineral	metal	tile	glass
std	*17.41*	*265.19*	*41.57*	*7.72*	*2.96*
LR	8.81	127.59	16.20	4.40	1.69
RFR	7.27	72.09	13.49	3.50	1.05
XGB	6.53	74.93	12.23	3.99	1.08

Since MAE is a metric relative to a sample, it should not be used for a direct comparison across samples. It is useful, however, to consider a significance of an error, per sample, relative to the sample's standard deviation (see Table 5). The smallest error achieved for *wood* was 6.53 t (standard deviation = 17.41 t), 72.09 t for *mineral* (std = 265.19 t), 12.23 t for *metal* (std = 41.57 t), and 1.05 t for *glass* (std=2.96 t).

The baseline LR model generalizes well on *wood* and *metal* samples (see Fig.3 left), but it suffers from an expected high bias error, indicating that the model is not able to learn sufficiently from the training data.

Prediction of *roof tile* quantities is considered unsuccessful for all tested algorithms, with R-squared value between 0.01 and 0.14 on the test set. Possible reasons and implications of specific results are discussed in the next section.

Discussion

Interpretation of results

The methodology delivers promising results for approximating the amounts of materials in buildings before demolition. The trained models can be applied to residential buildings and render a prediction, without the need for extensive documentation or in situ visits. The only information needed is a set of general building attributes, which can be queried from a local building register (e.g., GWR) or recognized from a building's image. The two investigated ensemble tree models (RFR and XGB) rendered very similar results, although RFR was more straightforward to train for a satisfactory learning skill. The XGB is a more complex algorithm than its counterpart, requiring more time and experience in the process of tuning hyperparameters. It is possible that in this study, XGB was too powerful for such a noisy dataset, which resulted in more pronounced overfitting. Having said that, the model's performance could possibly be further improved. This might be especially worthwhile when working with a bigger, quality-assured data sample.

The trends and differences between the two ensemble tree algorithms are in line with the experience of other researchers and practitioners (Mehta et al., 2019).

The dataset itself is considered the major limiting factor for the ML models' predictive skills. Insufficient training data is bound to compromise models' learning process. Indeed, the two smallest label sets (*roof tile* and *glass*) posed the most difficulties while training the models and rendered the weakest predictions. But even the bigger label sets could benefit from more and better-quality data. In addition to the amount of data, some materials had ambiguity in their input data classification, e.g., roof tile was sometimes reported within the general mineral waste category instead of in its own. This could further explain the poor prediction skill for these materials.

The class imbalance in categorical features (i.e., a significant variation in the number of instances per class) was acknowledged as a potential negative factor for ML performance, but its impact was not confirmed. It is possible that the negative impact was partially mitigated by using redundant features (two features describing location, and two features describing a building's age). In this proof of concept, the feature importance analysis and the impact of imbalanced or redundant features were not thoroughly analyzed and should be explored in detail in further studies.

A Data Architecture for Continuous Learning

Even though data availability was considered sufficient for the proof of concept, the results showed that further research and application of predictive models in this domain would highly benefit from a bigger, quality-assured dataset. To increase the prediction reliability, disposal proofs instead of disposal concepts could be used as data input. The former contains data on the material amounts reported after a demolition, while the latter only relies on pre-demolition audits. At the time of writing, the number of demolition proofs at UGZ constituted only roughly 5 percent of the overall building data, which is highly insufficient for ML model training. Nevertheless, data from disposal concepts alone could be collected and processed more automatically to save time and effort. In addition, a sample's representativeness over time is important if the goal is a continuous real-world application in the future. The current methodology only renders static models and is limited to making guesses on a 'frozen snapshot' of time. A continuously updated dataset would strongly improve a model's performance.

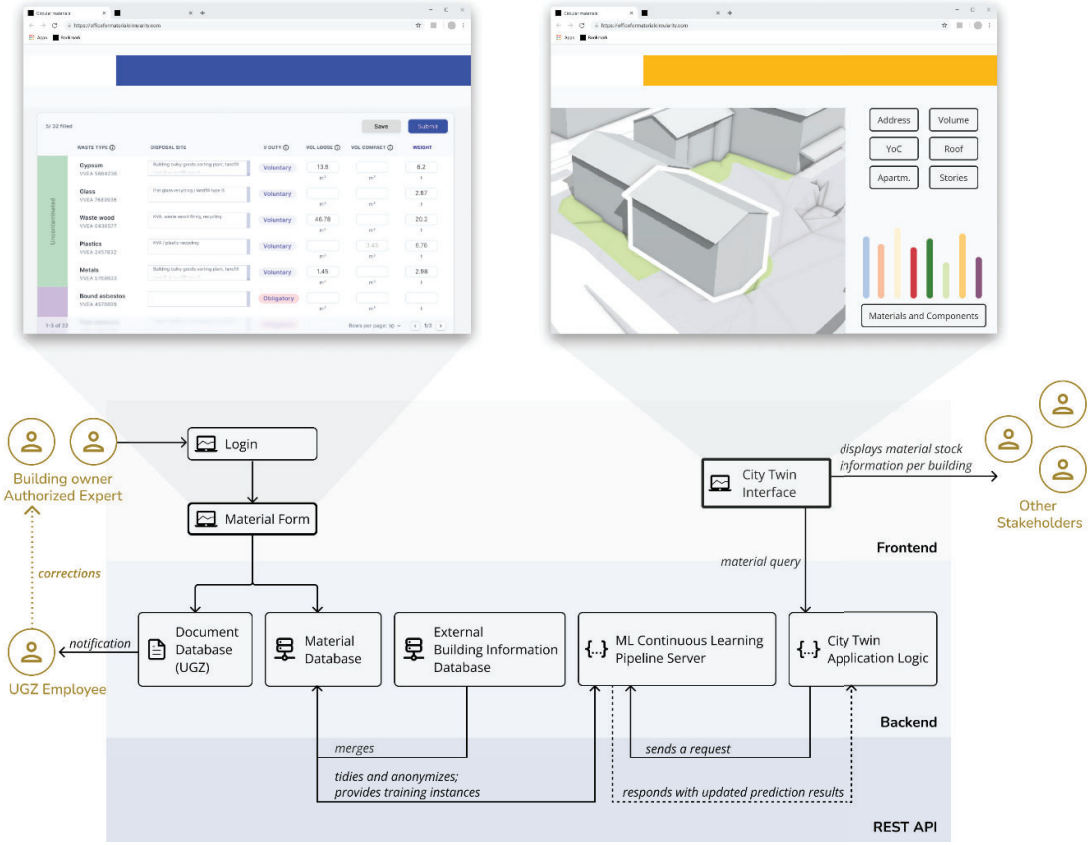

Figure 4: Bottom: A proposed concept of a framework. Data on material amounts in buildings is collected in a structured manner and then automatically processed and shared. Top Left: A mockup of a 'Material form' in a proposed online customer portal. Top Right: A mockup of a City Twin interface.

Therefore, we conceptualize a framework for improving the process of collecting, storing, and disseminating data on building materials in the existing buildings in Zurich (see Fig. 4). We consider it the next necessary step toward a dynamic ML model for the continuous prediction of material stock in buildings. A vast collection of quality-assured, consistent data would lay a foundation for further exploration of predictive algorithms, thus minimizing uncertainties stemming from the data quality. The proposed data architecture can be followed in Fig. 4 and is described in detail in the next paragraph.

The data pipeline starts as in the current process with the building owner submitting material information, supported by an authorized expert. Information about contaminated and non-contaminated materials would be recorded directly in the 'Material Form' in an online customer portal. A standardized form would help to eliminate the problem of multiple formats, ambiguous information, and missing documentation. For example, it would assure that materials are reported in their respective categories instead of being aggregated. For example, roof tile would need to be documented separately from the rest of mineral waste. Furthermore, building component information, useful for estimating reuse potential, could be requested (e.g., number of windows, sinks, doors, radiators). Adding this requirement could specifically help with estimating buildings component stocks, which is currently rarely present at the urban scale (Arora et al., 2019).

If such a form were filled out and submitted, data would flow to two separate databases. For the first, a .pdf would be sent to a UGZ employee for verification and conformity with current UGZ processes. The second database would store the information in a machine-readable format connected to related servers through a REST API client-server architecture. Material information could thus be merged for completeness with other building attributes accessed from external databases, e.g. GeoAdmin API, 2022. A continuous ML training pipeline would fetch the latest instances of complete data information to update the material stock forecasts. Stakeholders could then access and query up-to-date estimations of material amounts per building using a City Twin platform (such as LUUCY, 2022) for accessing other open-access building information.

The proposed framework is based on observed existing processes in the regulatory agencies and stakeholders' landscape of the City of Zurich. The automated data collection with an anonymization function addresses the problem of data inaccessibility due to privacy issues. Overall, the proposed process could expand the focus of the regulatory stakeholder from simply avoiding hazardous materials to supporting recovery of non-contaminated materials. Since similar demolition data is gathered throughout Switzerland (VVEA, 2020), other cities and authorities on a cantonal level could also benefit from the framework. Further research would need to specify the technical details of the framework and validate its applicability with different stakeholders. We expect that the global circularity movement in resource management would act as a motivating factor to embrace the proposed approach.

Conclusions

This paper showed the feasibility of applying a data-driven approach to material stock quantification in buildings, for available open and semi-open data in the City of Zurich. The amounts of five chosen material types can be predicted from the publicly available set of features. Both ensemble tree algorithms tested in this study exhibit a reasonable skill and strongly outperform the baseline LR model. Nevertheless, our findings show that only specific materials in a building stock could be predicted due to insufficient data. Even though it is important to further research and compare ML algorithms suitable for the investigated task, we therefore find it imperative to create reliable datasets first. To address this, we propose a new framework for the collection, processing, and dissemination of the data on buildings' materials and components. The framework relies on information already gathered by a city regulatory body and could modernize existing workflows by connecting public and private stakeholders. It would also benefit future researchers in their exploration of a broader spectrum of predictive algorithms in the domain. Although targeted to the context of the city of Zurich, other cities and municipalities could potentially adopt the framework to foster the circularity of construction materials and components at the urban scale.

Acknowledgments

We thank Patrick Buschor from UGZ Stadt Zürich for access to the documentation of demolition projects in Zurich.

References

Akanbi, L.A., Oyedele, A.O., Oyedele, L.O., Salami, R.O., 2020. Deep learning model for Demolition Waste Prediction in a circular economy. J. Clean. Prod. 274.

Akhtar, A., Sarmah, A.K., 2018. Construction and demolition waste generation and properties of recycled aggregate concrete: A global perspective. J. Clean. Prod. 186, 262–281.

Arora, M., Raspall, F., Cheah, L., Silva, A., 2019. Residential building material stocks and component-level circularity: The case of Singapore. J. Clean. Prod. 216, 239–248.

Baustoffrecycling Schweiz (2020) Verordnung über die Vermeidung und die Entsorgung von Abfällen (VVEA).

Bundesamt für Umwelt BAFU (2006) Richtlinie für die Verwertung mineralischer Bauabfälle.

Cha, G.W., Moon, H.J., Kim, Y.M., Hong, W.H., Hwang, J.H., Park, W.J., Kim, Y.C., 2020. Development of a prediction model for demolition waste generation

using a random forest algorithm based on small datasets. Int. J. Environ. Res. Public. Health 17, 1–15.

Entsorgungskonzept Rück- und Umbau (2020) Kanton Zürich. Available at: https:// ww.zh.ch/de/planen-bauen/baubewilligung/umgang-mitbauabfaellen/entsorgungskonzeptrueck-umbau.html (Accessed: 12.12.2022).

European Commission, 2016. EU Construction & Demolition Waste Management Protocol.

GeoAdmin API (2022) 3.1.0 documentation. Available at: https://api3.geo.admin.ch/index.html (Accessed: 8.12.2022).

geocat.ch (2018) Katalog Available at: https:// www.geocat.ch/geonetwork/srv/ger/catalog.search#/metadata/edf24f57-9b2f-8a15-b793-aefdb69a079d (Accessed: 12.12.22).

Heeren, N., Hellweg, S., 2018. Tracking Construction Material over Space and Time Prospective and Geo-referenced Modeling of Building Stocks and Construction Material Flows.

Honic, M., Hinterleitner, A., Schlögel, I., Kovacic, I., Sreckovic, M., 2021. Application of GPR-technology for identifying the material composition of building components. pp. 366–372.

KBOB, K. der B.L. der öffentlichen B., 2016. Ökobilanzdaten im Baubereich [WWW Document]. Available at: https://www.kbob.admin.ch/kbob/-de/home/themen-leistungen/nachhaltiges-bauen/-oekobilanzdaten_baubereich.html (Accessed 8 December 2022).

LUUCY (2022) Die Plattform für Raum- und Immobilienentwicklung. Available at: https:// www.luucy.ch/ (Accessed: 8.12.2022).

Mehta, P., Bukov, M., Wang, C.-H., Day, A.G.R., Richardson, C., Fisher, C.K., Schwab, D.J., 2019. A high-bias, low-variance introduction to Machine Learning for physicists. Phys. Rep., A high-bias, low-variance introduction to Machine Learning for physicists 810, 1–124.

Ostermeyer, Y., Nägeli, C., Heeren, N., Wallbaum, H., 2018. Building Inventory and Refurbishment Scenario Database Development for Switzerland. J. Ind. Ecol. 22, 629–642.

Stadt Zürich (2022) Erneuerung, Umbau, Abbruch. Available at: https://www.stadt-zuerich.ch/prd/de/-index/statistik/themen/bauen-wohnen/erneuerung/-erneuerung-umbau-abbruch.html (Accessed: 12.12.2022).

TABULA WebTool (2015) Building Typology - Available at: https://webtool.building-typology.eu-/#bm (Accessed: 12.8.22).

UN Environment and International Energy Agency. (2017) Towards a zero-emission, efficient, and resilient buildings and construction sector. Global Status Report 2017

Wu, P.Y., Sandels, C., Mjörnell, K., Mangold, M., Johansson, T., 2022. Predicting the presence of hazardous materials in buildings using machine learning. Build. Environ. 213.