Resources, Conservation & Recycling 199 (2023) 107253

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.elsevier.com/locate/resconrec

Machine learning models for the prediction of polychlorinated biphenyls and asbestos materials in buildings

Pei-Yu Wu [*,a,b], Claes Sandels [a], Tim Johansson [a], Mikael Mangold [a], Kristina Mjörnell [a,b]

[a] RISE Research Institutes of Sweden, 412 58 Gothenburg, Sweden
[b] Department of Building and Environmental Technology, Faculty of Engineering, Lund University, 221 00 Lund, Sweden

ARTICLE INFO

Keywords:
Hazardous material
Pre-demolition audits
Machine learning
Prediction
Probability distribution
Building stock

ABSTRACT

Hazardous materials in buildings cause project uncertainty concerning schedule and cost estimation, and hinder material recovery in renovation and demolition. The study aims to identify patterns and extent of polychlorinated biphenyls (PCBs) and asbestos materials in the Swedish building stock to assess their potential presence in pre-demolition audits. Statistics and machine learning pipelines were generated for four PCB and twelve asbestos components based on environmental inventories. The models succeeded in predicting most hazardous materials in residential buildings with a minimum average performance of 0.79, and 0.78 for some hazardous components in non-residential buildings. By employing the leader models to regional building registers, the probability of hazardous materials was estimated for non-inspected building stocks. The geospatial distribution of buildings prone to contamination was further predicted for Stockholm public housing to demonstrate the models' application. The research outcomes contribute to a cost-efficient data-driven approach to evaluating comprehensive hazardous materials in existing buildings.

1. Introduction

The potential presence of hazardous materials creates uncertainty for time and cost estimates in renovation and demolition projects as well as health concerns for workers (Powell et al., 2015; Rašković et al., 2020). A considerable quantity and variety of contaminants remain in the built environment nowadays, and their extent and location in buildings are rarely known due to insufficient documentation and verification (Bergsdal et al., 2014; Franzblau et al., 2020; Govorko et al., 2017; Wilk et al., 2019). As such, pre-demolition audit practice is introduced to identify the presence and quantify the amounts of hazardous materials to guide demolition contractors and waste handling companies in evaluating the contamination risk in material sorting (ECORYS, 2016; Wahlström et al., 2019b, 2019a). The audit inventories are also crucial for implementing selective demolition and deconstruction, as well as for quality assessment of reclaimed materials as a foundation for designing closed-loop circular strategies (Bergmans et al., 2017; Wahlström et al., 2020, 2019a). Nevertheless, the current time-consuming and costly practice of identifying hazardous materials on a building basis can hardly be used to estimate the remaining hazardous materials for the entire existing building stock. In addition, the quality and the completeness of inventory vary significantly between building types and regions, depending on the experience of auditors and building complexity (Wu et al., 2021a, 2021b). To prepare for and improve in-situ inspection of material quality appraisal and prevent second contamination in material recovery, exploring new approaches for enhancing the efficiency of in-situ hazardous material screening is necessary.

Since the 1990s, polychlorinated biphenyls (PCBs) have been used as impregnation agents in building components for enhancing fire resistance and electrical insulation, while asbestos was commonly used for sound and heat insulation as additives to glue, joints, paints, or plaster in construction. Several data-driven methods, such as material flow and stock analysis, were investigated to estimate residual PCB and asbestos stock for a longer timeframe at an aggregated level based on PCB field sampling and source-centric inventories, and the import and consumption data of asbestos products from demolition activities. However, studies of forecasting upcoming hazardous waste streams from a top-down perspective (Donovan and Pickin, 2016; Zoraja et al., 2021) or mapping local emission sources from a bottom-up approach (Bergsdal et al., 2014; Diamond et al., 2010; Diefenbacher et al., 2016; Shanahan et al., 2015) have limitations. The former concerns the certainty of

* Corresponding author at: Sven Hultins Plats 5, 412 58 Gothenburg, Sweden.
E-mail address: pei-yu.wu@ri.se (P.-Y. Wu).

https://doi.org/10.1016/j.resconrec.2023.107253
Received 30 June 2023; Received in revised form 18 September 2023; Accepted 9 October 2023

determining the lifetime of components and variation in sensitivity analysis due to extensive data sources and multiple assumptions, while the latter is fairly resource-demanding for a broad uptake in an urban-wide application.

In light of these shortcomings, machine learning is a rather cost-efficient, accurate, and scalable approach to computing multivariate prediction models. It offers an opportunity for exploiting past inventory records to identify the presence patterns of hazardous materials in buildings (Wu, 2022; Wu et al., 2022b, 2021c). Several prediction applications were developed in former studies and promising results were attained for mapping asbestos-containing products and roofing materials (Abriha et al., 2018; Krówczyńska et al., 2020; Wilk et al., 2019), estimating the probability distribution of radioactive concrete (Wu et al., 2023), and classifying the potential presence of asbestos and PCB-containing building materials (Wu et al., 2022b, 2022a). The results of these pilot studies may be used to improve in-situ hazardous material identification for intervention planning, such as source elimination and safe waste management.

This study deepened the analyses in the direction of expanding the prediction scope to multiple asbestos and PCB-containing materials and compared model robustness between regional building stocks. The primary focus was to identify critical variables and their synergies correlating with the presence of asbestos and PCB in buildings built between 1930 and 1985, and further on, attempting to quantify the extent of contaminated materials in the Swedish building stock. The chosen period is deemed suitable as substantial hazardous materials are frequently detected in buildings built before the ban on asbestos and PCB in the mid-70s in Sweden (Wu, 2022; Wu et al., 2021a). The research outcomes contribute to bridging the interdisciplinary areas between the EU Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH, EC 1907/2006) and the EU Resource efficiency opportunities in the building sectors legislations (COM 445, 2014), which can facilitate relevant authorities to formulate hazardous substances risk management policies for circular building (Swedish National Board of Housing Building and Planning, 2023). Other actors in the construction and demolition waste management sector, such as property owners, demolition contractors, and waste handling companies, could also benefit from risk-informed inspection planning. To realize the overarching research goals, three objectives are specified:

(i) To evaluate the quality and quantity of pre-demolition audit inventories and describe the statistics of PCB and asbestos-containing material detection in regional building stocks.
(ii) To develop machine learning models for predicting PCB and asbestos-containing materials and analyze prediction outcomes.
(iii) To estimate and visualize the probability distribution of PCB and asbestos-containing materials in regional building stocks.

2. Material and methods

The study was designed in three parts: (i) data sources assembling, (ii) data preprocessing and validation, and (iii) machine learning modeling and prediction, illustrated in Fig. 1.

2.1. Data sources assembling

Pre-demolition audit inventories conducted between 2010 and 2022 were collected from the city archives of the most populated Swedish municipalities – Stockholm, Gothenburg, Malmö, as well as Kiruna, a smaller town situated in the north where many buildings were demolished due to mining activities. After data cleaning, i.e., removal of buildings lacking registers and buildings categorized as complementary buildings, etc., 788 observations remained, of which 47% were from Stockholm, 35% from Gothenburg, 10% from Malmö, and 8% from Kiruna. This data sampling had a broad geographical coverage representing heterogeneous building stock and diverse inventory types, including consultancy reports, protocol, control plans, and demolition plans. Then the inventories were digitalized into a machine-readable dataset, where the purposes and the extent of the investigation, inspection date, auditors, decontamination, and detection of hazardous materials were recorded following the Resource and waste guidelines for construction and demolition (Byggföretagen, 2019) to ensure high data quality and documentation coherence. The dataset contained 69% of buildings with total inspection and 14% of buildings with decontamination records of PCB or/and asbestos.

Meanwhile, the building registers from the Swedish Energy

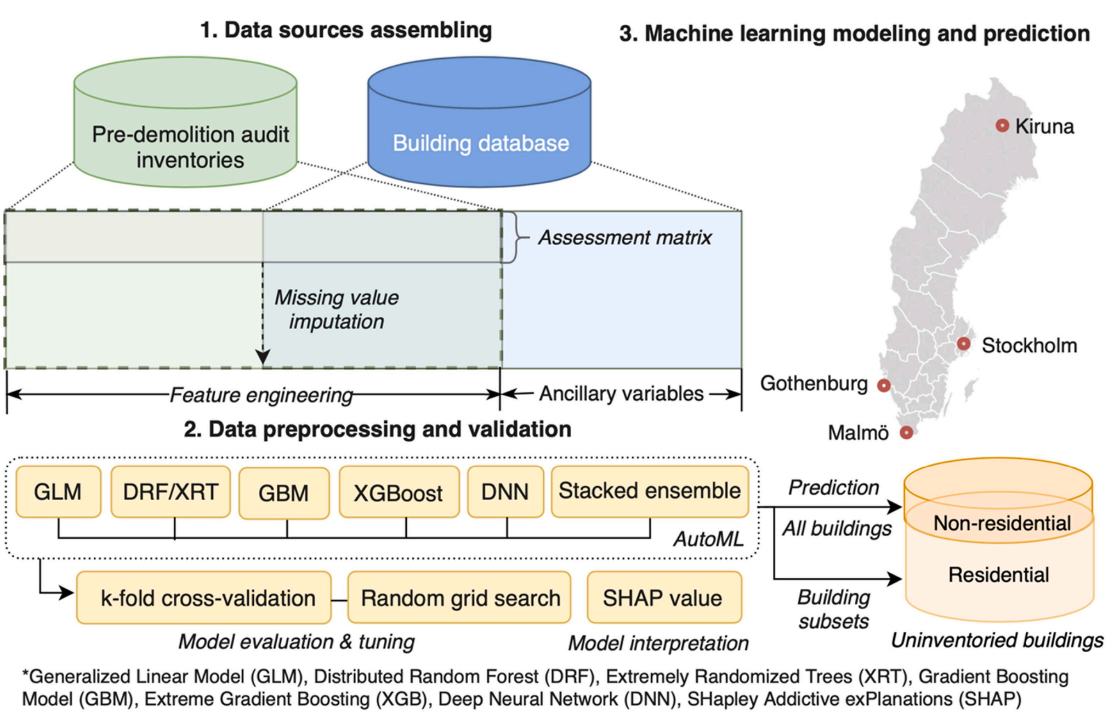

Fig. 1. Study outline.

Performance Certificates (EPCs), the Swedish Real Estate Taxation register, and the Municipal Cadastral register, were compiled and merged via the Feature Manipulation Engine (FME) to supplement generic information, including cadastral affiliation, building usage and characteristics (Johansson et al., 2017). Coupling multiple registers was beneficial in comparing the data quality between database updates and enhanced dataset completeness. Then the data merging uncertainties were assessed by characterizing the matching relationships of the number of properties and buildings. To evaluate the representativeness of the training dataset to the prediction dataset, the mean and missing values of features were examined against the building stock of the sampled municipalities in Appendix A Table A1. The building registers of the observed buildings were used as features for model training, while the registers of uninspected buildings were employed as the input data for prediction.

2.2. Data preprocessing and validation

The data preprocessing concerning data evaluation, missing data imputation, and feature engineering, was applied to the dataset integrated from inventories and building registers. Based on the building use type code, data were stratified into various building classes – residential (i.e., single-family and multifamily houses) and non-residential building (i.e., school, commercial/office, and industrial buildings) subsets to obtain a rather homogeneous observation cluster. Despite non-residential buildings only accounting for a small part of the Swedish building stock, its sample proportion in the environmental inventories was rather large since many of them were built before the 80s. Specific building classes, i.e., public and school buildings, have more quality inventory data as they are renovated frequently with decontamination records. To maximize data utilization and compare patterns of asbestos and PCB components in non-residential buildings with literature, a study scope on comprehensive hazardous materials for the entire building stock was determined. The statistical description of the features between the training and prediction dataset was further described in Appendix A Table A2. The quality and quantity of inventories were assessed using Eq. (1) to determine potential hazardous materials for modeling, referred to the previous work by Wu et al. (2021a). The weighting scheme for different inventory types was configured based on the detailed level of hazardous material information and the experience level of auditors. Consultancy reports and protocols contained detection records of hazardous components, whereas control and demolition plans only indicated the presence of hazardous substances.

$$y = \frac{(Ir \times nr + Ip \times np + Ic \times nc + Id \times nd)}{N} \times K \tag{1}$$

y = Assessment score [0–100].
I = Inventory type for weighting individual observations. $I = 1$ if it is a report (r), $I = 0.75$ if it is a protocol (p), $I = 0.5$ if it is a control plan (c), and $I = 0.25$ if it is a demolition plan (d).
n = The number of observations in the studied subgroup $[0 < n]$.
N = The number of observations in the entire dataset.
K = Weight based on data size. $K = 1$ if $n >= 400$, $K = 0.75$ if $300 =< n < 400$, $K = 0.5$ if $200 =< n < 300$, $K = 0.25$ if $100 =< n < 200$, $K = 0$ if $n < 100$.

Missing values of predictive variables were computed and replaced through the *k*-nearest neighbors algorithm (k-NN) using the mean value from the two nearest neighbors found in the training set for data size and quality improvement. The raw features comprised geographical attributes, i.e., postcode, cadastral attributes, i.e., building category and types, and building parameters, i.e., construction year, heated area, the number of floors, basements, stairwells, apartments, and ventilation types. Then the derived features, such as building physical footprint, area per apartment, and stairwell, were created for modeling.

2.3. Machine learning model and prediction

The clean dataset was partitioned into an 80% training set and a 20% validation set taking account of sample weights of target variables through Python scikit-learn and H_2O libraries, which provide distributed and scalable machine learning and predictive analytics. Automated machine learning (AutoML) automates steps in the machine learning pipeline using a wrapper function to train various algorithms for multiple machine learning tasks simultaneously and providing explainability for groups of models or individual models, including hyperparameter tuning and model evaluation (Ledell and Poirier, 2020). Then the class imbalance between binary labels and materials for each data subgroup was addressed by stratifying and specifying balance classes in the model configuration. Six classifiers were trained and compared via the AutoML pipeline, including the Distributed Random Forest (DRF, including Extremely Randomized Trees XRT), Generalized Linear Model with regularization (GLM), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), fully-connected multi--layer Deep Neural Network (DNN), and Stacked Ensembles (stacking a meta-learner at the second-level to find the optimal combination of base learners). In the two-fold prediction, 20 sub-models for all buildings set, and ten sub-models for residential and non-residential building sets were created. Various parameter settings were experimented with, including variable encoding, feature combination, and appending sample weights, and then extensive training and fine-tuning were carried out by random grid search given the small data size, and the training stopped until the pre-defined maximum training time was reached. Thereafter, the performance of models was ranked on the leaderboard based on logarithmic loss (error rates), and the lead models from the 5-fold cross-validation and the prediction of the validation set were selected according to the highest F1 (weighted average of Precision and Recall calculated by predicted class), and the optimal area under the receiver operating characteristic curve (ROC AUC, true positive rate against false positive rate calculated by predicted probability value). The feature impacts and the impact magnitude of the prediction outcomes from lead models were analyzed through methods such as confusion matrix, variable importance, partial dependence plots, and SHapley Additive exPlanations (SHAP) summary (Rodríguez-Pérez and Bajorath, 2020).

Subsequently, the lead models were applied to the regional building databases to predict PCB and asbestos-containing materials in buildings without pre-demolition audit inventories, of which the potential presence of PCB and asbestos in Stockholm, Gothenburg, Malmö, and Kiruna municipalities were estimated by predicted class and compared their probability distribution by construction year among all buildings and building type subsets. Furthermore, the probability of containing specific PCB and asbestos materials was predicted in a showcase for the municipality-owned public housing companies in Stockholm considering the high performance of models for residential buildings and the large sample size of the training set from Stockholm. Using the organization number in the taxation registers, 1910 residential buildings with complete registers were retrieved from building owners Familjebostäder, Stockholmshem, and Svenska Bostäder with a building composition of 1878 multifamily houses, 20 rowhouses, seven attached houses, and five chain houses, and supplied to the models of residential buildings for prediction. Then the prediction outcomes of three PCB-containing materials and eight asbestos-containing materials were visualized on the building footprint maps, which were derived from the OpenStreetMap, using Python geopandas, osmnx (Boeing, 2017), and plotly Mapbox libraries for geospatial analysis.

3. Results and discussion

The section presents the results in four parts – data analytics of the inventories, predictive model development via machine learning, model output interpretation and visualization, and probability estimation of PCB and asbestos-containing materials in existing buildings.

3.1. Data analytics

Explorative data analysis was performed based on the collected inventories to attain an overview of the presence of hazardous materials in buildings and determine data quality and quantity for modeling. The detection rate and the data size of hazardous materials in each municipality are illustrated in Appendix A Table A3. The average detection of PCB and asbestos in the selected regional building stocks was 47% and 78%. A closer analysis showed that 43 % of observations contained both asbestos and PCB, 36% contained either of the substances, and only 21% of observations had neither substance. Approximately 8% of buildings contained more than two types of PCB materials, while the corresponding percentage of asbestos materials was 47%. Among all, PCB capacitors in lights or burners were found in more than half (51%) of the buildings, and around one-fifth of the buildings contained PCB joints or sealed double-glazing windows. On the other hand, the most frequently detected asbestos materials were pipe insulation (65%), door or window insulation (61%), and cement panels (60%), followed by floor mats (49%) and joints (49%), ventilation channel (42%), and carpet glue (40%). Only one-third of the buildings were found with asbestos tile or clinker. Overall, the patterns of PCB joints and sealed double glazing windows, as well as asbestos valves, panels, tile or clinker, and joints or sealants were alike in Stockholm and Gothenburg with minor variation. Malmö and Kiruna had relatively few inspection records and thus the uncertainty of detection rates was high.

Furthermore, the inventory data were applied to the equation described in Section 2.2 to identify hazardous materials with large numbers of detailed detection records. The data assessment matrix in Appendix A Fig. A1 presents the assessment scores of PCB and asbestos-containing materials computed by Eq. (1) across building classes and municipalities. All types of buildings except single-family houses in Stockholm and Gothenburg had fair data quality and quantity. The inventories from Malmö only contained non-residential buildings with low data quality and size, and Kiruna had insufficient and less comprehensive inventories with several blank areas (no records) and zero scores (less than five observations). According to the indication of the matrix scores and detection frequency, four PCB materials (i.e., joints, windows, capacitors, acrylic flooring) and eight asbestos materials (i.e., pipe insulation, door or window insulation, panel, tile or clinker, carpet glue, floor mat, ventilation channel, joints) were determined as the target variables for prediction.

The detection frequency and the types of hazardous materials varied between building typologies, depending on material choice and the construction tradition of the area where buildings were situated. Complex buildings such as multifamily houses, schools, commercial buildings, and industrial buildings tended to contain multiple PCBs (more than three types) or asbestos materials (more than five types), particularly those built between 1955 and 1980. Despite some of them having recent renovation years, the extent and the scope of renovation work could hardly be known from either registers or inventory, thus it could not be used as a parameter to determine the likelihood of hazardous material detection. Moreover, the detection rates may slightly be overestimated due to the potential data selection bias from renovated or demolished buildings. The information of whether materials “existed” or “were not detected” in buildings was unavailable in current inventories, therefore, creating uncertainties in statistics. Former studies by Franzblau et al. (2020), Govorko et al. (2019), and Song et al. (2016) computed statistics of asbestos material found in residential buildings in the U.S. and Australia, and public buildings in Korea, yet the prevalent asbestos materials and their detection rates were incomparable to this study as the building stock is country dependent. A similar conclusion applies to the result comparison of PCB quantification with Diamond et al. (2010) in Canada and Diefenbacher et al. (2016) in Switzerland.

3.2. Predictive model development

In the predictive model development, 36 training iterations were initiated for twelve hazardous materials on three datasets, and the best performance of each algorithm was benchmarked according to the leaderboards, presented in Table 1. Details on the leader models concerning algorithm types and hyperparameters, data size and class distribution, maximized F1 threshold and error rates for train and validation sets can be found in Appendix B Table B1. The findings showed that machine learning classifiers attained promising results in predicting asbestos door or window insulation (AUC_{AB} and $F1_{AB}$: 0.85, AUC_{R} and $F1_{R}$: 0.94, AUC_{nR} and $F1_{nR}$: 0.84), asbestos pipe insulation (AUC_{AB} and $F1_{AB}$: 0.82, AUC_{R} and $F1_{R}$: 0.85, AUC_{nR} and $F1_{nR}$: 0.79), and PCB capacitors (AUC_{AB} and $F1_{AB}$: 0.82, AUC_{R} and $F1_{R}$: 0.85, AUC_{nR} and $F1_{nR}$: 0.79). Asbestos ventilation channel (AUC_{AB} and $F1_{AB}$: 0.76, AUC_{R} and $F1_{R}$: 0.89) and joints (AUC_{AB} and $F1_{AB}$: 0.78, AUC_{R} and $F1_{R}$: 0.83) could be predicted well in all buildings and residential building sets. The presence patterns of asbestos panels (AUC_{R} and $F1_{R}$: 0.87), tile or clinker (AUC_{R} and $F1_{R}$: 0.81), and carpet glue (AUC_{R} and $F1_{R}$: 0.87) in residential buildings were also identified. However, the PCB acrylic floor in the residential building set was highly imbalanced and failed in training because of one cardinality (the number of possible values that a feature can assume). On the other hand, the models trained on the non-residential set performed equally or slightly worse than those trained on all types of buildings, which was assumed to be due to the heterogeneous building constitutions. For the rest of the materials, the prediction rates were lower and the deviations between AUC and F1 were substantial.

The DRF/XRT, GBM, XGBoost, and stacked ensemble models appeared to be more suitable for predicting asbestos and PCB materials. The GLM models seemed to be too simple to capture the material presence patterns, while DNN models were too complicated for training in such small datasets. The results of model performance aligned approximately with previous studies by Wu et al. (2022b) that reported an average of 78 % recall rate for asbestos pipe insulation in multifamily houses and 83 % recall rate for PCB joint in school buildings using logistic regression, SVM, k-NN, and tree ensembled classifiers. Applying artificial neural networks to identify asbestos materials in residential dwellings indicated similar prediction performance – pipe insulation (AUC: 0.80), door and window insulation (AUC: 0.70), panel (AUC: 0.65), tile or clinker (AUC: 0.51), carpet glue (AUC: 0.73), floor mats (AUC: 0.42), ventilation channel (AUC: 0.86) (Wu et al., 2022a). To further minimize false negative or false positive prediction, other less common metrics for cost-sensitive learning can be considered in the future, such as the Gini coefficient (inequality quantification among values of a frequency distribution based on the Lorenz curve for measuring the quality of a binary classifier), Absolute MCC (correlation indicators for the actual and predicted values by setting the threshold for the model’s confusion matrix to a value that generates the highest Matthews Correlation Coefficient), F2 (extra weight adjustment for recall rates to penalize models with more false negatives than false positives).

Data quality and quantity were the primary limitations of the study. Given the well-trained models had reached the optimal bias and variance trade-off, the irreducible errors could be attributed to uncertainties in data matching derived from data quality, for instance, 20% of observations had insufficient information in either inventory or registers for data merging, 3% observations merged with duplicated register or parts of parameters were unmatched, and 1% observations were coupled between registers and aggregated inventory. Several hindrances related to data size were also encountered in modeling highly sparse and class-imbalanced datasets of various regions, building classes, and hazardous materials. Large amounts of missing data, i.e., the numbers of stairwells and apartments data led to the loss of extra features. In addition, critical features relevant to the prediction task might be lacking and result in difficulty in label distinguishment, such as the information on the location of historical asbestos products manufacture plants, suggested

Table 1
Performance evaluation of average AUC and F1 score from cross-validation and validation set of the machine learning models for prediction of PCB and asbestos-containing materials (ACM). The values in bold signified leader models for respective hazardous material prediction.

Algorithm	GLM		DRF/XRT		GBM		XGBoost		DNN		Stacked ensemble	
Metrics (e-2)	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1
All buildings												
PCB joint	64	42	76	56	**78**	**57**	74	54	66	46	–	–
PCB windows	–	–	–	–	–	–	**60**	**37**	–	–	60	37
PCB capacitors	73	77	**82**	**79**	**82**	**79**	79	77	80	79	–	–
PCB acrylic floor	24	7	41	12	61	14	48	11	44	12	**66**	**16**
ACM pipe insul*	65	81	79	83	**80**	**84**	78	85	73	82	–	–
ACM door insul*	74	80	85	85	**84**	**86**	83	85	80	83	–	–
ACM panel	68	78	**71**	**79**	70	80	66	78	**71**	**79**	–	–
ACM tile/clinker	57	54	73	63	**75**	**65**	73	63	65	59	–	–
ACM carpet glue	64	62	73	66	**75**	**68**	71	65	68	64	–	–
ACM floor mat	–	–	–	–	–	–	**63**	**67**	–	–	62	68
ACM ventilation**	–	–	79	72	78	70	76	70	75	68	**80**	**71**
ACM joint/sealant	70	71	77	74	**80**	**76**	75	74	69	70	–	–
Residential buildings (incl. single-family and multifamily houses)												
PCB joint	84	67	81	61	**88**	**69**	81	58	–	–	85	68
PCB windows	56	32	–	–	–	–	**61**	**35**	–	–	46	22
PCB capacitors	**86**	**79**	84	75	82	74	66	62	–	–	79	73
PCB acrylic floor	–	–	–	–	–	–	–	–	–	–	–	–
ACM pipe insul*	72	87	**81**	**88**	74	88	75	90	–	–	78	87
ACM door insul*	73	90	88	94	**93**	**94**	84	92	–	–	90	94
ACM panel	76	87	**84**	**90**	77	86	75	87	–	–	82	88
ACM tile/clinker	55	67	74	70	**83**	**78**	82	75	–	–	80	73
ACM carpet glue	69	71	–	–	–	–	**88**	**85**	–	–	88	83
ACM floor mat	46	76	–	–	–	–	**61**	**76**	–	–	54	76
ACM ventilation**	84	85	89	86	85	84	83	83	–	–	**90**	**87**
ACM joint/sealant	–	–	**78**	**87**	79	86	74	87	–	–	75	86
Non-residential buildings (incl. school, office/commercial, and industrial buildings)												
PCB joint	58	41	–	–	–	–	**75**	**53**	–	–	72	52
PCB windows	**66**	**47**	68	44	60	37	66	42	–	–	66	43
PCB capacitors	75	79	–	–	–	–	**76**	**80**	–	–	75	78
PCB acrylic floor	41	10	–	–	–	–	**68**	**18**	–	–	37	6
ACM pipe insul*	65	79	76	81	**75**	**82**	73	81	–	–	74	81
ACM door insul*	80	82	–	–	–	–	**85**	**83**	–	–	**85**	**83**
ACM panel	57	73	**65**	**77**	62	77	59	76	–	–	**65**	**77**
ACM tile/clinker	60	53	67	57	**69**	**61**	63	53	–	–	66	56
ACM carpet glue	70	65	72	66	**73**	**67**	71	65	–	–	69	67
ACM floor mat	54	64	67	66	**68**	**66**	65	65	–	–	67	67
ACM ventilation**	70	73	**74**	**65**	68	63	66	62	–	–	72	64
ACM joint/sealant	61	63	72	66	**76**	**71**	73	68	–	–	73	67

* Asbestos-containing material pipe insulation, asbestos-containing material door or window insulation.
** Asbestos-containing material ventilation channel.

correlated to asbestos used by Wilk et al. (2019), as well as the change rates of components related to renovation history and ownership types.

3.3. Model interpretation

To decode the gray box models, the variable importance heatmap across models for hazardous materials prediction for all buildings was compiled in Appendix B Fig. B1. Features, such as construction year, floor area, building physical footprint, and postcode, the number of floors, were commonly recognized as crucial for several hazardous material predictions. After that, the leader tree models of three hazardous materials with the highest performance— PCB capacitors, asbestos pipe insulation, and asbestos door and window insulation — were chosen for output interpretation in residential and non-residential building subsets. In the SHAP summary plots, feature values of the observations were normalized at the y-axis and configured based on their SHAP values at the x-axis, indicating the impact magnitude of each predictive variable in descending order. Fig. 2 shows that the patterns of hazardous materials vary between regional building stocks. Residential buildings with large floor areas and equipped with exhaust ventilation as well as non-residential buildings, specifically office/commercial, and industrial buildings, built in the earlier ages without balanced ventilation with heat exchangers were more likely to contain PCB capacitors. Similar patterns were found in asbestos door and window insulation with considerable feature impact magnitude, where post-war residential buildings with large floor areas and building physical footprint, and old non-residential buildings (except for school buildings) with basements were vulnerable. The feature value directions were less evident in asbestos pipe insulation prediction which might be attributed to pipe changing, and features associated with residential buildings were floor area, the number of floors, building physical footprint, construction year, as well as postcode, construction year, floor area, and number of floors for non-residential buildings.

The global interpretations of aggregated SHAP values were in good agreement with the literature. Diamond et al. (2010) and Shanahan et al. (2015) pointed out that PCB capacitors and joints were installed extensively in proportion to the building volume and electrical demand. Song et al. (2016) emphasized that the likelihood of detecting asbestos materials increased significantly by building age and building physical footprint. The earlier work from the authors also showed that these two variables were strongly correlated to the presence of asbestos pipe insulation in multifamily houses. The local explanation of individual observations of specific feature impact values could be obtained through the query at the either index or dataset level.

3.4. Probability estimation of PCB and asbestos in regional building stocks

The pre-trained models of the three selected PCB and asbestos materials were employed in the building registers of four municipalities to estimate predicted labels for the three datasets on the regional scale, as

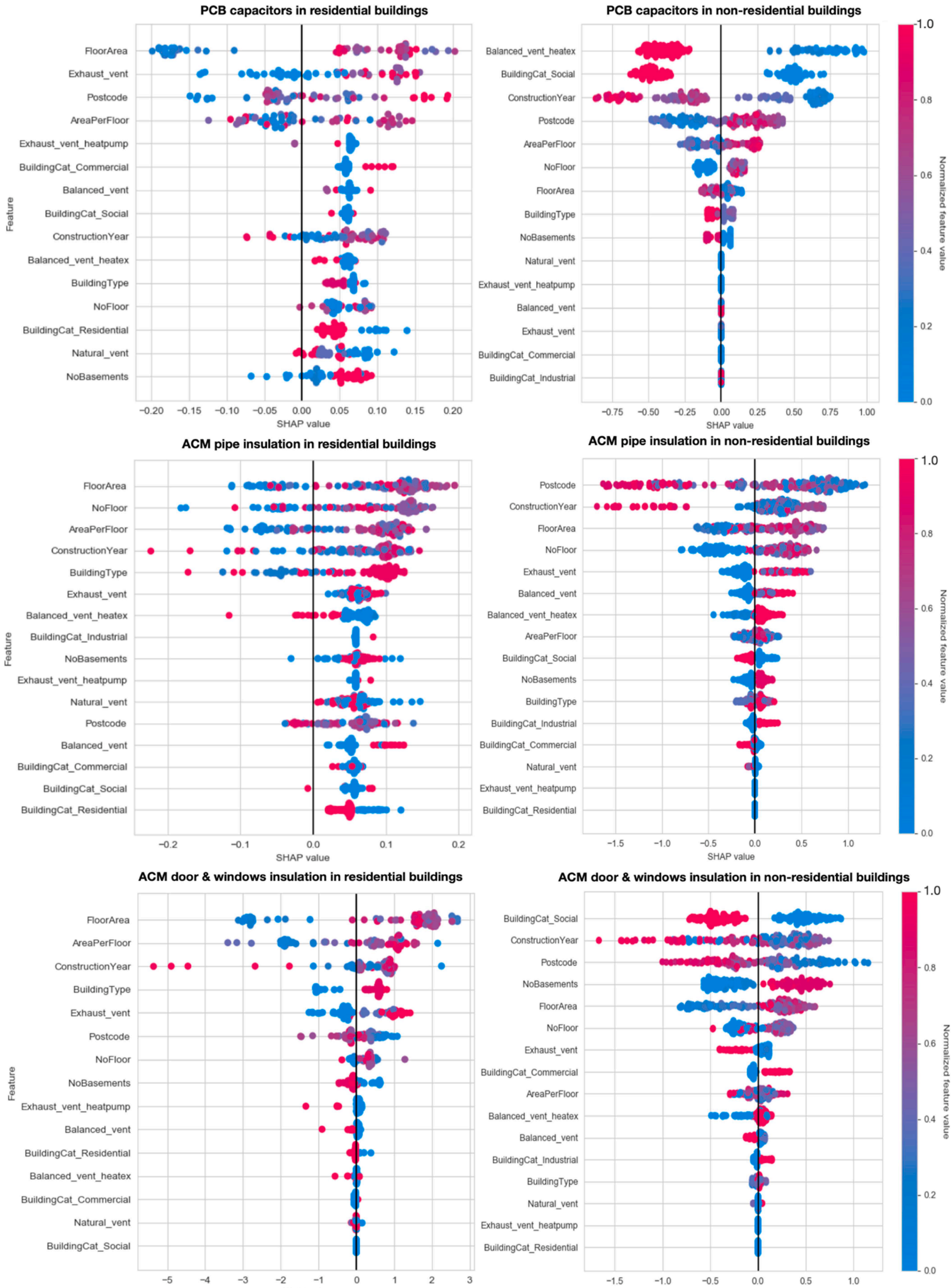

Fig. 2. Feature impact and magnitude based on the SHAP values in residential and non-residential building subsets: (i) PCB capacitors, (ii) asbestos pipe insulation, (iii) asbestos door and window insulation. The postcode initial in Stockholm is 1, Malmö 2, Gothenburg 4, and Kiruna 9. The building type code for school buildings is 8–21, single-family houses are 30–32 and 35, multifamily houses are 33, industrial buildings are 40–53, and office and commercial buildings are 99.

shown in Table 2. For regional residential building stock, it was estimated that around 2%, 57%, and 14% of buildings contained PCB capacitors, asbestos pipe insulation, and door and window insulation. On the contrary, the corresponding shares for non-residential buildings were 44%, 92%, and 60%. The findings complied with the existing expert assumptions that these hazardous materials were more pervasive in non-residential buildings. The average predicted detection rates in all buildings were lower than the previous statistics in Appendix A Table A3, which reflected the actual building stock constitution of 95% of residential buildings in comparison to the residential building shares in inventories of 29% for PCB capacitors, 37% for asbestos pipe insulation and 33% for door and window insulation.

The probability distribution of hazardous materials in regional buildings was further investigated by municipalities and building types across building ages. The results in Fig. 3 agreed with the previous statistics of the predicted class and exhibited interesting trends. Compared to non-residential buildings, residential buildings had lower probabilities of containing PCB capacitors and asbestos door and window insulation, except for the residential buildings in Kiruna. However, the tendency reversed in the mid-70s when the probability of asbestos pipe insulation in non-residential buildings dropped significantly, and the downward trends could also be observed in the other two hazardous materials that were assumed due to the PCB and asbestos bans in Sweden. Yet surprisingly, the probability of PCB capacitor and asbestos pipe insulation in residential buildings did not decrease over time compared to the trend identified in the asbestos door and window insulation. Besides, the confidence intervals for the residential buildings in Stockholm, Gothenburg, and Malmö were smaller than non-residential buildings, while the Kiruna data showed divergent development and were regarded as invalid and unrepresentative unless more training data from the municipality was collected. The overall patterns at the building stock level offered insights to relevant authorities and municipalities for formulating decontamination policies and guidelines for the existing buildings.

The choropleth maps in Fig. 4 display the predicted geospatial

Table 2
Estimating the share of buildings potentially containing PCB and asbestos materials in the regional building stocks.

Non-inventoried buildings	Stockholm	Gothenburg	Malmö	Kiruna	Total
Residential buildings (*N*)	39,248	39,225	16,725	146	95,344
Non-residential buildings (*N*)	1563	1324	885	16	3788
All buildings (*N*)	40,811	40,549	17,610	162	99,132
PCB capacitor					
%Residential (AUC: 0.86, F1: 0.79)	0.01	0.04	0.01	0.78	0.02
%Non-residential (AUC: 0.76, F1: 0.80)	0.30	0.58	0.48	0.56	0.44
%All buildings (AUC: 0.82, F1: 0.79)	0.12	0.19	0.28	0.99	0.18
ACM pipe insulation					
%Residential (AUC: 0.81, F1: 0.88)	0.63	0.43	0.77	0.31	0.57
%Non-residential (AUC: 0.75, F1: 0.82)	0.98	0.83	0.98	0.75	0.92
%All buildings (AUC: 0.80, F1: 0.84)	0.56	0.28	0.56	0.81	0.44
ACM door and window insulation					
%Residential (AUC: 0.91, F1: 0.94)	0.20	0.10	0.12	0.12	0.14
%Non-residential (AUC: 0.85, F1: 0.83)	0.57	0.51	0.80	0.25	0.60
%All buildings (AUC: 0.84, F1: 0.86)	0.23	0.14	0.23	0.21	0.19

distribution of hazardous materials for a subset of municipality-owned residential buildings in Stockholm. The probability indication at the building level allowed building owners to assess contamination exposure based on the probability cluster, offered decision-support for risk-based inspection planning, and suggested detailed material sampling before renovation and (selective) demolition. To apply the data-driven application on the national scale, it is necessary to include more inventories from other municipalities for training to reduce misclassification, meanwhile, taking local building stock differences into account. By introducing diverse observations, the sample representativeness of the geographical coverage will improve (Clemmensen and Kjærsgaard, 2022) with a sufficient sample size for constructing individual models to overcome building class imbalance between sampling and population. The checkpointing models delivered in the study lay a foundation for the future when more data are available and can be integrated into web visualization interfaces.

Other statistical and machine learning modeling applications were developed to facilitate in-situ hazardous material identification and source separation (Wu et al., 2021c). Recognition of asbestos cement roofing on the macro scale was made possible by applying convolutional neural networks (Krówczyńska et al., 2020; Raczko et al., 2022) and random forest algorithms (Abriha et al., 2018; Wilk et al., 2019) on aerial imagery, field inventories, and building databases, while detection of asbestos fiber on the micro-scale was succeeded through principal component analysis with hyperspectral imagery (Bonifazi et al., 2019, 2018). To characterize multiple asbestos materials, descriptive statistics at the building level were obtained based on input data from mobile app surveys (Govorko et al., 2019), waste audit inspection reports (Franzblau et al., 2020), and asbestos product databases (Mecharnia et al., 2019). However, the results of these methods had limited predictive capacity and high estimation uncertainties. To overcome the shortcomings, the granularity and diversity of prediction outcomes were improved by constructing models from data with temporal and geographical representativeness. Compared to previous studies by the authors (Wu et al., 2023, 2022b, 2021a), the novelty of the paper was highlighted by demonstrating the machine learning pipeline from model training, interpretation, prediction to outcome utilization. As the models were developed from building registers and pre-demolition audit inventories accessible in many European countries, they have promising potential for methodological replication and upscaling.

4. Conclusions

The study systematically investigated the presence of hazardous materials in the built environment and ascertained the prediction potential of machine learning using pre-demolition audit inventories as input data. Beyond describing statistical results for individual contaminants like former research, the proposed predictive approach in the study enables a comprehensive and cost-efficient evaluation of multiple in-situ hazardous materials for existing buildings. PCB and asbestos were found in 47% and 78% of inventoried buildings, of which 43% of observations contained both substances and 36% contained either of the substances. The data assessment matrix showed that multifamily houses, school buildings, offices/commercial and industrial buildings contained sufficient data labels with satisfactory quality for modeling. The leader models varied between prediction tasks, but the Distributed Random Forest (including Extremely Randomized Trees), Gradient Boosting Machine, Extreme Gradient Boosting, and Stacked Ensembles models usually attained optimal performance. Nearly all PCB and asbestos materials could be predicted with average AUC and F1 above 0.79 in the residential building set except for PCB windows and acrylic floor and asbestos floor mat.

Key building features related to the presence of hazardous materials were determined and the extent of potentially contaminated building stock was estimated by leveraging the developed models. Components

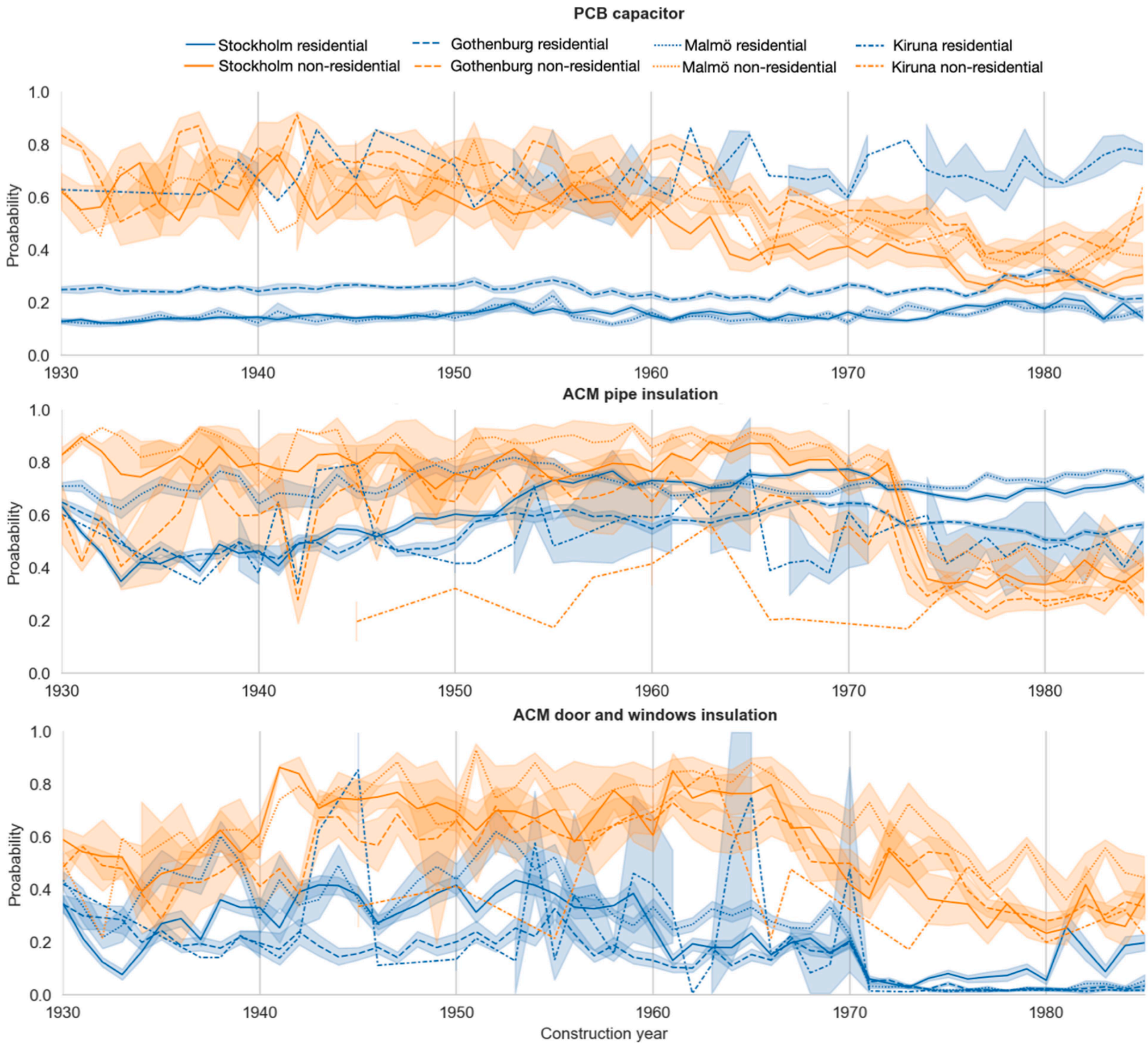

Fig. 3. Probability distribution of PCB-containing capacitors, asbestos-containing pipe insulation, and door and window insulation in regional building stocks.

with PCB and asbestos materials were more common in post-war buildings with large floor areas and building physical footprints. In addition, building category, postcode, and ventilation types were also indicators useful in predicting hazardous materials. It was estimated that 18%, 44%, and 19% of regional buildings contain PCB capacitors, asbestos pipe insulation, and door and window insulation respectively, which were situated approximately within the statistical range considering varied building type composition between the actual and the inventoried building stocks. Detailed probability distributions over the construction year showed distinctive patterns and certainty between residential and non-residential buildings with minor variations between regions, providing a diagnostic overview of asbestos and PCB-prone building stock for municipalities and the housing authority to implement the EU Construction and Demolition Waste Management Protocol (ECORYS, 2016). The geospatial probability visualization of in-situ PCB and asbestos in the Stockholm public housing pinpointed high- risk buildings, which could be useful for property owners to make informed decisions when planning for pre-demolition audit. Future research is suggested to include more inventories from different municipalities and refine the machine learning models on the building typological and regional basis for better generalizability.

Funding

The research leading to these results has received funding from the Horizon 2020 Research and Invention Programme, under Grant Agreement No 957026, the BuiltHub project, the Swedish Foundation for Strategic Research (SSF) with grant number FID18-0021, and the Re: Source project from the Swedish Energy Agency with grant number P2022-00304.

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

CRediT authorship contribution statement

Pei-Yu Wu: Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft, Writing – review & editing,

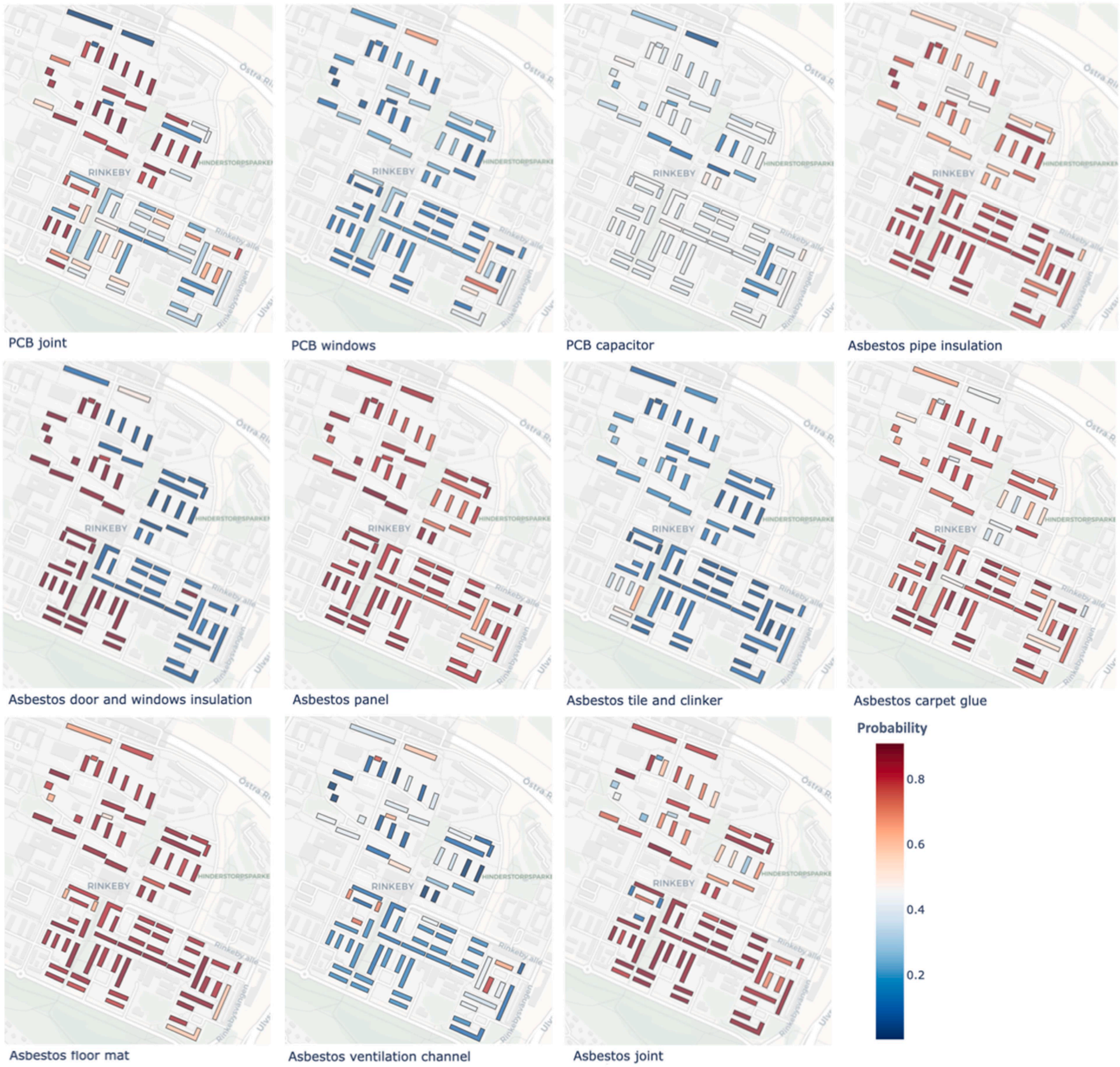

Fig. 4. Visualizing the geospatial probability distribution of asbestos and PCB materials in the Stockholm public housing buildings.

Visualization. **Claes Sandels:** Supervision, Data curation, Methodology, Writing – review & editing. **Tim Johansson:** Data curation, Writing – review & editing. **Mikael Mangold:** Supervision, Funding acquisition, Writing – review & editing, Project administration. **:** Supervision, Funding acquisition, Writing – review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

The research work is part of the PhD project "Prediction of Hazardous Materials in Buildings using Machine Learning" and is related to the Re: Source project "Development of machine learning-embedded applications to support hazardous material assessment in renovation and demolition". Special thanks are sent to Frida Palstam, Olivia Heuts, and Birgitta Gunér for assisting digitalization of pre-demolition audits.

Appendix A Data analytics and feature selection

Tables A1, A2, A3, Fig. A1

Table A1
The statistical description of the features between the training and prediction dataset.

Category	Feature	Feature representation	Train		Prediction	
			Mean	%Missing	Mean	%Missing
		Raw feature				
Geographics	Postcode	[1–, 2–, 4–, 9DDDD]	–	0	–	0
Cadastral affiliation	Building category	[1–4]	–	0	–	0
	Building type	[1–99]	–	0	–	0
Building parameter	Construction year	[1930–1985]	1960	0	1960	0
	Floor area (m^2)	[–]	4337	10	797	20
	Numbers of floors	[–]	3	16	3	70
	Number of basements	[0,1,2,3+]	1	26	1	69
	Number of stairwells	[1–]	3	47	2	71
	Number of apartments	[–]	31	50	22	71
	Exhaust ventilation	[0,1]	–	43	–	52
	Balanced ventilation	[0,1]	–	43	–	52
	Balanced heat exchanger	[0,1]	–	43	–	52
	Exhaust with heat pump	[0,1]	–	44	–	52
	Natural ventilation	[0,1]	–	43	–	52
		Derived feature				
Building parameter	Building physical footprint (m^2)	[–]	935	21	595	70
	Area per stairwell (m^2)	[0–]	1846	67	1287	83
	Area per apartment (m^2)	[0–]	151	69	117	76

Table A2
Sample distribution of the training and prediction datasets.

Subset	Building class	Category*	Type**	%Train (N)	%Prediction (N)
Residential building	Single-family house	1	30,31,32,35	18 (138)	80 (83,079)
	Multifamily house	1	33	19 (149)	16(16,598)
Non-residential building	School building	3	8,19,21	24 (190)	1 (1333)
	Office/commercial	4	99	20 (159)	2 (2475)
	Industrial building	2	40–53	19 (150)	1 (973)
Total				100 (786)	100 (104,458)

* Building category according to property registers: 1- Residential, 2- Industry, 3- Society function, 4- Workplace.
** Building type according to property registers: 8- College, 19- School, 21- University, 30- Detached house, 31- Detached chain house, 32- Semi-detached houses, 33- Multifamily building, 35- Small house with several apartments, 40- Other manufacturing industry, 41- Gas turbine plant, 42- Industrial hotel, 43- Chemical Industry, 44- Condensation power plant, 45- Nuclear power plant, 46- Food industry, 47- Metal or machine industry, 48- Textile industry, 49- Wood products industry, 50- Hydropower plant, 51- Wind turbine, 52- Heating plant, 53- Other industrial building, 99- Unspecified.

Table A3
The detection rate and data size of hazardous materials in each municipality.

Municipality	Stockholm		Gothenburg		Malmö		Kiruna		Total	
Count	368		276		80		64		786	
	%rate	N	%rate	N	%rate	N	%rate	N	%rate	N
PCB	43	256	51	227	37	67	71	21	47	571
Joint/sealant	22	187	18	157	22	59	33	3	21	406
Sealed windows	15	169	18	133	17	29	40	5	17	336
Capacitors	49	118	50	147	17	12	93	14	51	291
Acrylic flooring	4	139	3	116	0	15	0	1	3	271
Door closer	36	55	56	18	50	4	100	1	42	78
Cable with oil	11	71	24	21	50	2	100	1	16	95
Others*	100	2	43	30	47	15	100	1	48	48
Asbestos	78	344	75	254	84	76	88	25	78	699
Pipe insulation	71	254	48	160	84	37	94	17	65	468
Valves	30	66	35	72	80	5	100	1	35	144
Door/win insulation**	67	197	46	142	87	23	90	10	61	372
Panel	58	172	52	103	78	32	94	17	60	322
Tile/clinker	33	229	36	174	46	39	62	13	36	455
Carpet glue	45	187	26	139	68	41	44	9	40	376
Floor mat	53	182	39	131	67	46	33	6	49	365
Vent. channel	23	163	58	116	65	20	100	11	42	310
Switchboard	83	6	16	61	100	3	100	1	27	71
Joint/sealant	48	144	42	64	54	46	86	7	49	261
Others*	60	114	73	63	42	31	100	14	64	222

* Other asbestos-containing materials are for example, gasket, roofing felt, expansion vessel, spray insulation, etc., while other PCB materials are floor color, oil in elevator machine, contaminated concrete, etc.
** Door or window insulation.

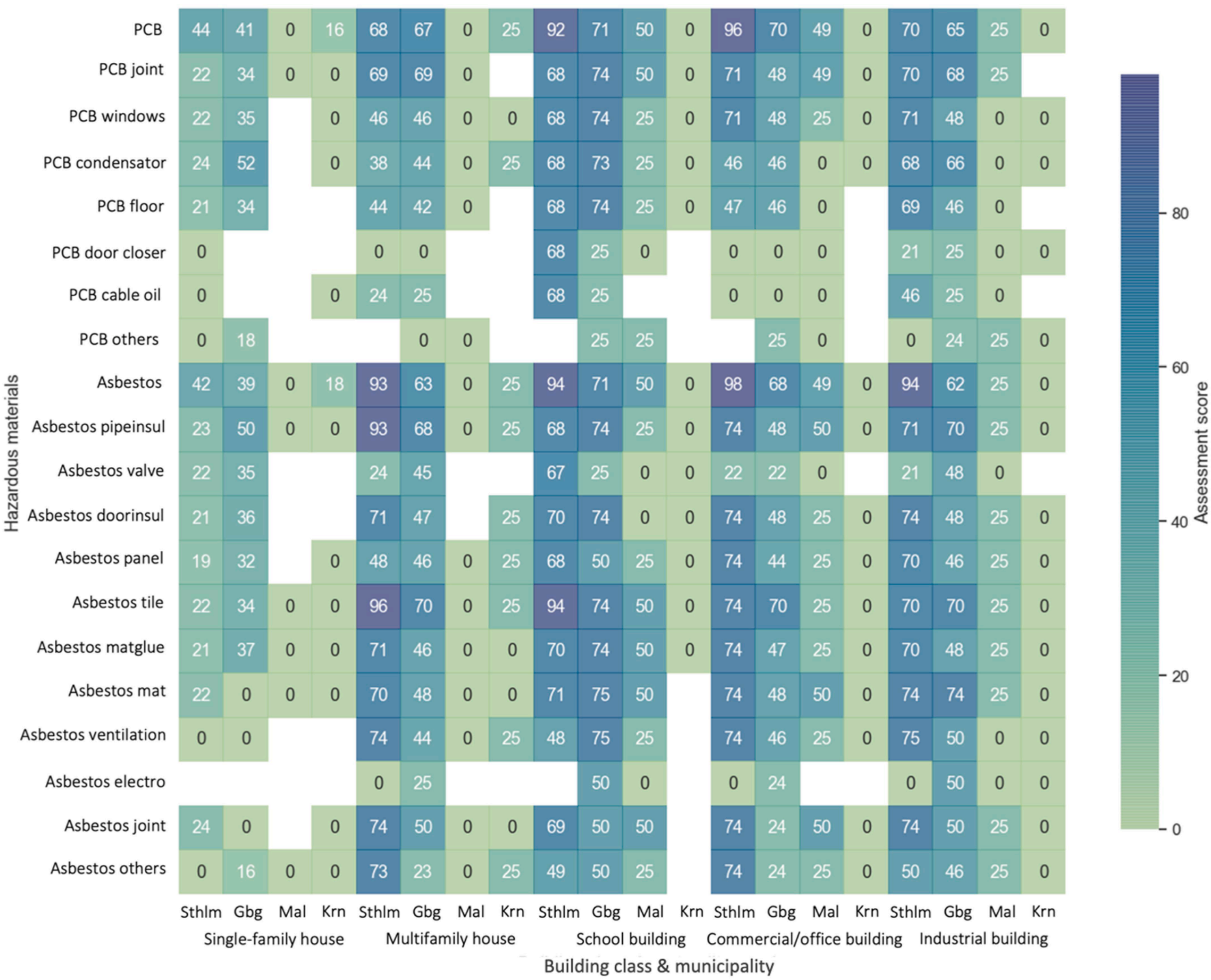

Fig. A1. Overview of the data assessment matrix by assigning quantile range weights for the data quality according to the inventory sources and the data size threshold of a minimum of five. The x-labels refer to Stockholm (Sthlm), Gothenburg (Gbg), Malmö (Mal), and Kiruna (Krn).

Appendix B Predictive model development and interpretation

Table B1, Fig. B1.

Table B1
Overview of the lead machine learning models. StackEn entails stacked ensemble models.

	Data size	Detection		Leader model	Max F1 threshold		Error rate		Hyperparameter
		0	1		Train	Val	Train	Val	
All buildings									
PCB joint	406	321	85	GBM	0.19	0.27	0.21	0.16	No. trees: 24 No. depths: 6 No. leaves: 26–56
PCB windows	336	280	56	XGBoost	0.13	0.06	0.25	0.50	No. trees: 34
PCB capacitors	291	144	147	GBM	0.34	0.43	0.30	0.19	No. trees: 24 No. depths: 6 No. leaves: 25–45
PCB acrylic floor	271	262	9	StackEn	0.04	0.01	0.27	0.65	No. bases: 3 Baselearner: GBM, DNN Metalearner: GLM
ACM pipe insulation	468	163	305	GBM	0.41	0.62	0.22	0.23	No. trees: 28 No. depths: 6 No. leaves: 32–57
ACM door or window insulation	372	146	226	GBM	0.56	0.53	0.18	0.19	No. trees: 37 No. depths: 5–9 No. leaves: 12–18

(continued on next page)

Table B1 (*continued*)

	Data size	Detection		Leader model	Max F1 threshold		Error rate		Hyperparameter
		0	1		Train	Val	Train	Val	
ACM panel	322	129	193	XRT	0.48	0.48	0.30	0.28	No. trees: 26 No. depths: 12–20 No. leaves: 46–86
ACM tile/clinker	455	291	164	GBM	0.15	0.32	0.41	0.25	No. trees: 21 No. depths: 6 No. leaves: 29–50
ACM carpet glue	376	224	152	GBM	0.36	0.37	0.29	0.32	No. trees: 27 No. depths: 5–9 No. leaves: 13–17
ACM floor mat	365	185	180	XGBoost	0.38	0.43	0.45	0.41	No. trees: 30
ACM ventilation channel	310	181	129	StackEn	0.43	0.51	0.27	0.21	No. bases: 3 Baselearner: GBM, DRF, DNN Metalearner: GLM
ACM joint	261	134	127	GBM	0.35	0.41	0.28	0.26	No. trees: 36 No. depths: 4–7 No. leaves: 8–11
Residential buildings									
PCB joint	104	84	20	GBM	0.03	0.01	0.16	0.19	No. trees: 31 No. depths: 6 No. leaves: 7–34
PCB windows	86	75	11	XGBoost	0.02	0.50	0.63	0.17	No. trees: 32
PCB capacitors	83	53	30	GLM	0.49	0.60	0.17	0.12	Family: binomial Regular.: Ridge Lambda: 0.1001
PCB acrylic floor	75	73	2	–	–	–	–	–	–
ACM pipe insulation	172	48	124	DRF	0.48	0.62	0.18	0.17	No. trees: 39 No. depths: 6–12 No. leaves: 23–36
ACM door or window insulation	123	41	82	GBM	0.14	0.89	0.10	0.08	No. trees: 56 No. depths: 6 No. leaves: 10–30
ACM panel	101	30	71	DRF	0.54	0.94	0.20	0.10	No. trees: 21 No. depths: 6–13 No. leaves: 16–29
ACM tile/clinker	147	80	67	GBM	0.14	0.22	0.35	0.17	No. trees: 23 No. depths: 6 No. leaves: 25–45
ACM carpet glue	108	60	48	XGBoost	0.45	0.55	0.19	0.10	No. trees: 36
ACM floor mat	94	39	55	XGBoost	0.09	0.11	0.40	0.37	No. trees: 32
ACM ventilation channel	96	44	52	StackEn	0.36	0.13	0.16	0.15	No. bases: 6 Baselearner: GBM, DRF, GLM Metalearner: GLM
ACM joint	70	23	47	DRF	0.54	0.62	0.25	0.14	No. trees: 32 No. depths: 5–11 No. leaves: 9–19
Non-residential buildings									
PCB joint	302	237	65	XGBoost	0.26	0.61	0.32	0.20	No. trees: 33
PCB windows	250	205	45	GLM	0.21	0.23	0.26	0.14	Family: binomial Regular.: Ridge Lambda: 0.242
PCB capacitors	208	91	117	XGBoost	0.42	0.53	0.31	0.19	No. trees: 32
PCB acrylic floor	196	189	7	XGBoost	0.07	0.03	0.10	0.30	No. trees: 56
ACM pipe insulation	296	115	181	GBM	0.58	0.21	0.22	0.28	No. trees: 27 No. depths: 6–7 No. leaves: 14–20
ACM door or window insulation	249	105	144	XGBoost	0.43	0.49	0.30	0.12	No. trees: 31
ACM panel	221	99	122	StackEn	0.38	0.32	0.28	0.38	No. base: 3 Baselearner: GBM, DRF Metalearner: GLM
ACM tile/clinker	308	211	97	GBM	0.18	0.37	0.36	0.27	No. trees: 22 No. depths: 6–8 No. leaves: 11–22
ACM carpet glue	268	164	104	GBM	0.21	0.29	0.39	0.30	No. trees: 20 No. depths: 6–7 No. leaves: 14–17
ACM floor mat	271	146	125	GBM	0.26	0.40	0.38	0.35	No. trees: 28 No. depths: 6–8 No. leaves: 13–17
ACM ventilation channel	214	137	77	XRT	0.34	0.28	0.29	0.26	No. trees: 33 No. depths: 9–16 No. leaves: 18–58
ACM joint	191	111	80	GBM	0.20	0.28	0.36	0.26	No. trees: 26 No. depths: 5–7 No. leaves: 10–13

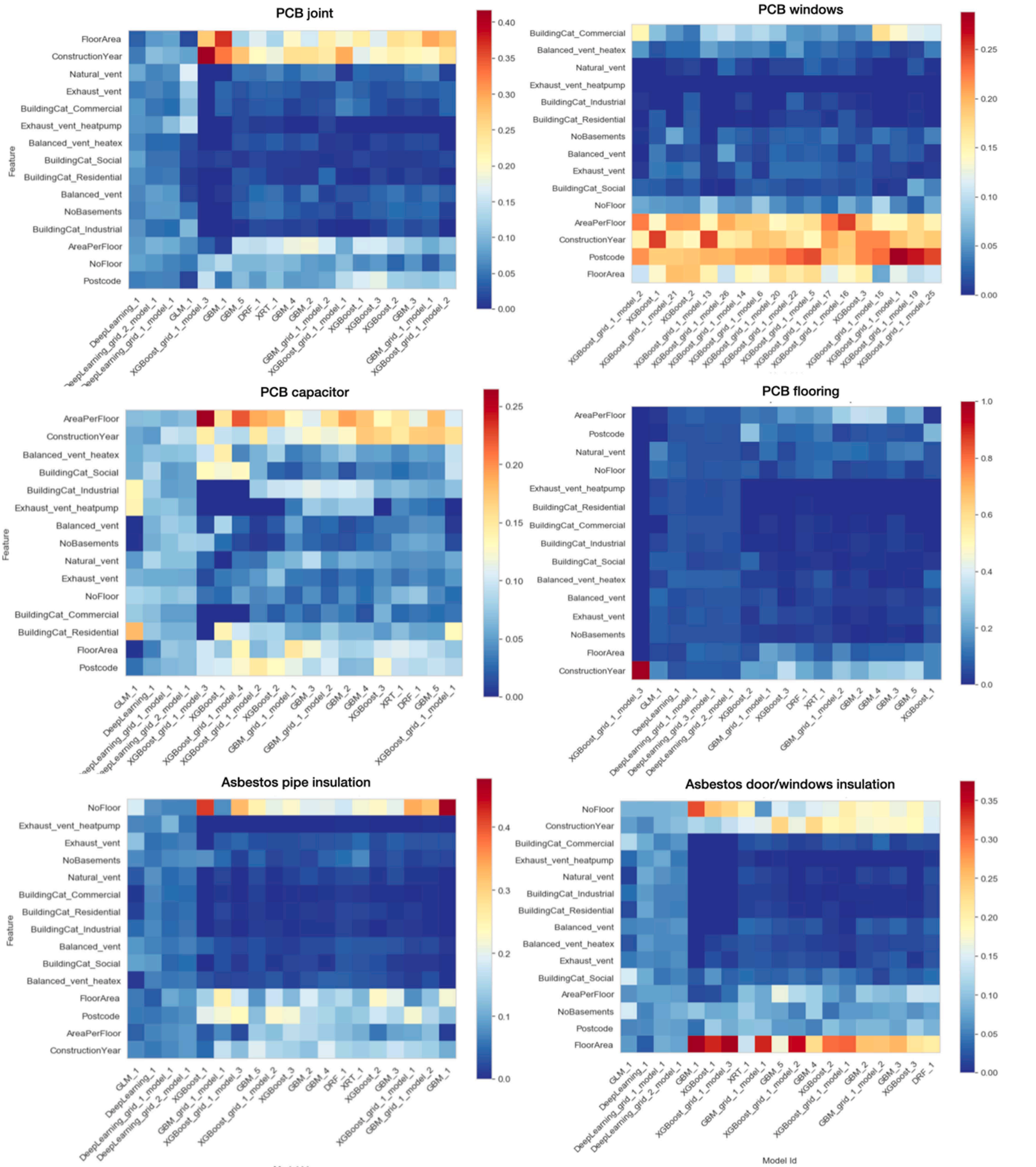

Fig. B1. Variable importance for asbestos and PCB materials in models for all buildings.

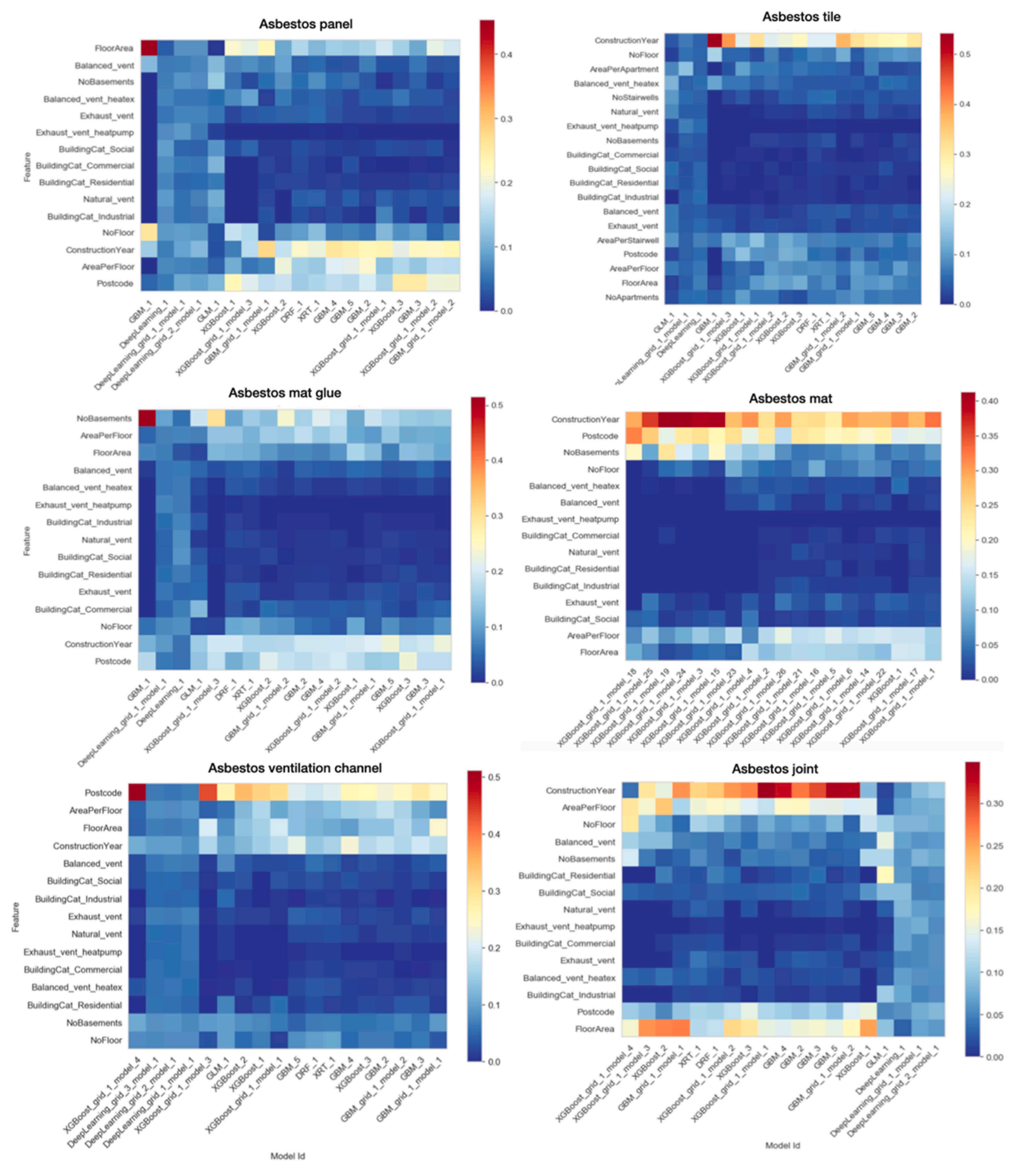

Fig. B1. (*continued*).

References

Abriha, D., Kovács, Z., Ninsawat, S., Bertalan, L., Balázs, B., Szabó, S., 2018. Identification of roofing materials with discriminant function analysis and random forest classifiers on pan-sharpened worldview-2 imagery – a comparison. Hung. Geogr. Bull. 67, 375–392. https://doi.org/10.15201/hungeobull.67.4.6.

Bergmans, J., Dierckx, P., Broos, K., 2017. Semi-selective demolition : current demolition practices in Flanders. In: HISER Conference. Delft, The Netherlands. https://doi.org/10.5281/zenodo.817324.

Bergsdal, H., Brattebø, H., Müller, D.B., 2014. Dynamic material flow analysis for PCBs in the Norwegian building stock. Build. Res. Inf. 42, 359–370. https://doi.org/10.1080/09613218.2014.887898.

Boeing, G., 2017. OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. Comput. Environ. Urban Syst. 65, 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004.

Bonifazi, G., Capobianco, G., Serranti, S., 2019. Hyperspectral imaging and hierarchical PLS-DA applied to asbestos recognition in construction and demolition waste. Appl. Sci. (Switzerland) 9, 1–15. https://doi.org/10.3390/app9214587.

Bonifazi, G., Capobianco, G., Serranti, S., 2018. Asbestos containing materials detection and classification by the use of hyperspectral imaging. J. Hazard. Mater. 344, 981–993. https://doi.org/10.1016/j.jhazmat.2017.11.056.

Byggföretagen, 2019. Resource and waste guidelines for construction and demolition.

Clemmensen, L.H., Kjærsgaard, R.D., 2022. Data representativity for machine learning and AI systems. In: FAccT '22: ACM Conference on Fairness, Accountability and Transparency, June 21-24, 2022, pp. 1–16. Seoul, South Korea 1.

Diamond, M.L., Melymuk, L., Csiszar, S.A., Robson, M., 2010. Estimation of PCB stocks, emissions, and urban fate: will our policies reduce concentrations and exposure? Environ. Sci. Technol. 44, 2777–2783. https://doi.org/10.1021/es9012036.

Diefenbacher, P.S., Gerecke, A.C., Bogdal, C., Hungerbühler, K., 2016. Spatial distribution of atmospheric PCBs in Zurich, Switzerland: do joint sealants still matter? Environ. Sci. Technol. 50, 232–239. https://doi.org/10.1021/acs.est.5b04626.

Donovan, S., Pickin, J., 2016. An Australian stocks and flows model for asbestos. Waste Manag. Res. 34, 1081–1088. https://doi.org/10.1177/0734242×16659353.

ECORYS, 2016. EU construction & demolition waste management protocol. Brussels, Belgium.

Franzblau, A., Demond, A.H., Sayler, S.K., D'Arcy, H., Neitzel, R.L., 2020. Asbestos-containing materials in abandoned residential dwellings in Detroit. Sci. Total Enviro. 714, 136580 https://doi.org/10.1016/j.scitotenv.2020.136580.

Govorko, M., Fritschi, L., Reid, A., 2019. Using a mobile phone app to identify and assess remaining stocks of in situ asbestos in Australian residential settings. Int. J. Environ. Res. Public Health 16. https://doi.org/10.3390/ijerph16244922.

Govorko, M., Fritschi, L., White, J., Reid, A., 2017. Identifying asbestos-containing materials in homes: design and development of the ACM check mobile phone app. JMIR Form Res. 1 https://doi.org/10.2196/formative.8370 e7.

Johansson, T., Olofsson, T., Mangold, M., 2017. Development of an energy atlas for renovation of the multifamily building stock in Sweden. Appl. Energy 203, 723–736. https://doi.org/10.1016/j.apenergy.2017.06.027.

Krówczyńska, M., Raczko, E., Staniszewska, N., Wilk, E., 2020. Asbestos-cement roofing identification using remote sensing and convolutional neural networks (CNNs). Remote Sens. (Basel) 12, 1–16. https://doi.org/10.3390/rs12030408.

Ledell, E., Poirier, S., 2020. H_2O AutoML : scalable automatic machine learning. In: 7th ICML Workshop on Automated Machine Learning (AutoML).

Mecharnia, T., Khelifa, L.C., Pernelle, N., Hamdi, F., 2019. An approach toward a prediction of the presence of asbestos in buildings based on incomplete temporal descriptions of marketed products. In: K-CAP 2019 - Proceedings of the 10th International Conference on Knowledge Capture. Marina del Rey, United States, pp. 239–242. https://doi.org/10.1145/3360901.3364428.

Powell, J., Jain, P., Bigger, A., Townsend, T.G., 2015. Development and application of a framework to examine the occurrence of hazardous components in discarded construction and demolition debris: case study of asbestos-containing material and lead-based paint. J. Hazard. Toxic Radioact. Waste 19, 05015001. https://doi.org/10.1061/(asce)hz.2153-5515.0000266.

Raczko, E., Krówczyńska, M., Wilk, E., 2022. Asbestos roofing recognition by use of convolutional neural networks and high-resolution aerial imagery. Testing different scenarios. Build. Environ. 217 https://doi.org/10.1016/j.buildenv.2022.109092.

Rašković, M., Ragossnig, A.M., Kondracki, K., Ragossnig-Angst, M., 2020. Clean construction and demolition waste material cycles through optimised pre-demolition waste audit documentation: a review on building material assessment tools. Waste Manag. Res. 38, 923–941. https://doi.org/10.1177/0734242×20936763.

Rodríguez-Pérez, R., Bajorath, J., 2020. Interpretation of machine learning models using Shapley values: application to compound potency and multi-target activity predictions. J. Comput. Aided Mol. Des. 34, 1013–1026. https://doi.org/10.1007/s10822-020-00314-0.

Shanahan, C.E., Spak, S.N., Martinez, A., Hornbuckle, K.C., 2015. Inventory of PCBs in Chicago and opportunities for reduction in airborne emissions and human exposure. Environ. Sci. Technol. 49, 13878–13888. https://doi.org/10.1021/acs.est.5b00906.

Song, S.-J., Jang, B.-K., Jo, B.-H., Kim, Y.-J., Heo, E.-H., Lee, J.-D., Son, B.-S., Lee, J.-W., 2016. An asbestos risk assessment and areal distribution of asbestos containing materials in public buildings. J. Korean Soc. Occup. Environ. Hyg. 26, 267–276. https://doi.org/10.15269/jksoeh.2016.26.3.267.

Swedish National Board of Housing Building and Planning, 2023. Boverket ska hjälpa byggsektorn att utvecklas mot en cirkulär ekonomi [WWW Document]. URL https://www.regeringen.se/pressmeddelanden/2022/02/boverket-ska-hjalpa-byggsektorn-att-utvecklas-mot-en-cirkular-ekonomi/(accessed 2.23.22).

Wahlström, M., Bergmans, J., Teittinen, T., Bachér, J., Smeets, A., Paduart, A., 2020. Construction and Demolition Waste : Challenges and Opportunities in a Circular Economy. Eionet Report - ETC/WMGE 2020/1, Mol, Belgium.

Wahlström, M., Teittinen, T., Kaartinen, T., Liesbet, van C., 2019a. Hazardous Substances in Construction Products and Materials: PARADE. Best Practices for Pre-demolition Audits Ensuring High Quality RAw Materials. Esbo, Finland.

Wahlström, M., Zu Castell-Rüdenhausen, M., Hradil, P., Smith, K.H., Oberender, A., Ahlm, M., Götbring, J., Hansen, J.B., 2019b. Improving Quality of Construction & Demolition Waste-Requirements for Pre-Demolition Audit. Copenhagen, Denmark. https://doi.org/10.6027/TN2019-508.

Wilk, E., Krówczyńska, M., Zagajewski, B., 2019. Modelling the spatial distribution of asbestos-cement products in Poland with the use of the random forest algorithm. Sustainability (Switzerland) 11. https://doi.org/10.3390/su11164355.

Wu, P.-Y., 2022. Predicting Hazardous Materials in the Swedish Building Stock Using Data Mining. Lund University.

Wu, P.-Y., Johansson, T., Mangold, M., Sandels, C., Mj, K., 2023. Estimating the probability distributions of radioactive concrete in the building stock using Bayesian networks. Expert Syst. Appl. 222 https://doi.org/10.1016/j.eswa.2023.119812.

Wu, P.-Y., Mangold, M., Sandels, C., Johansson, T., Mjörnell, K., 2022a. Modeling artificial neural networks to predict asbestos-containing materials in residential buildings. In: IOP Conference Series: Earth and Environmental Science. https://doi.org/10.1088/1755-1315/1122/1/012050.

Wu, P.-Y., Mjörnell, K., Mangold, M., Sandels, C., Johansson, T., 2021a. A data-driven approach to assess the risk of encountering hazardous materials in the building stock based on environmental inventories. Sustainability (Switzerland) 13, 1–26.

Wu, P.-Y., Mjörnell, K., Mangold, M., Sandels, C., Johansson, T., 2021b. Tracing hazardous materials in registered records : a case study of demolished and renovated buildings in Gothenburg. J. Phys. Conf. Ser. 2069. https://doi.org/10.1088/1742-6596/2069/1/012234.

Wu, P.-Y., Mjörnell, K., Sandels, C., Mangold, M., 2021c. Machine learning in hazardous building material management: research status and applications. Recent Prog. Mater. 03 https://doi.org/10.21926/rpm.2102017, 1–1.

Wu, P.-Y., Sandels, C., Mjörnell, K., Mangold, M., Johansson, T., 2022b. Predicting the presence of hazardous materials in buildings using machine learning. Build. Environ. 213 https://doi.org/10.1016/j.buildenv.2022.108894.

Zoraja, B., Ubavin, D., Stanisavljevic, N., Vujovic, S., Mucenski, V., Hadzistevic, M., Bjelica, M., 2021. Assessment of asbestos and asbestos waste quantity in the built environment of transition country. Waste Manag. Res. https://doi.org/10.1177/0734242×211064031.