Chair of Computing in Civil and Building Engineering
TUM School of Engineering and Design
Technical University of Munich

IFC Model Analysis to gain Insights into Design Processes

Software Lab 2025

Submitted by

Meng-Ju Hsieh	03797997
Tan-An Ha	03799839
Kai Dietmair	03742268

Submitted on: December 17, 2025

Abstract

This project demonstrates a method for analysing Industry Foundation Classes (IFC) models by combining Stacking Ensemble Learning with a Human-in-the-Loop (HITL) mechanism. While machine learning models achieved high accuracy in initial tests (RMSE 0.154 for thermal transmittance), their reliability is constrained by the lack of diverse, real-world training data. To address this, an interactive dashboard serves as a bridge between the inference engine and the human expert. The system automatically parses uploaded IFC files, detecting anomalies such as inconsistent U-values or classification errors. Users can inspect these flagged objects and provide binary validation or corrective input. Validated data is subsequently fed back into the system, expanding the knowledge base and refining future predictions.

Contents

Acronyms **1**

1 Introduction **2**
1.1 Motivation . . . 2
1.2 Scope of this Report . . . 2

2 State of the Art **3**
2.1 Semantic Enrichment . . . 3
2.2 Human-in-the-Loop (HITL) . . . 3
2.3 Ensemble Learning . . . 5
2.4 Research Gap . . . 5

3 Methodology and Framework **6**
3.1 General Framework and Data Overview . . . 6
3.1.1 System Architecture . . . 6
3.2 Exploratory Data Analysis (EDA) . . . 7
3.2.1 Dataset Composition and Acquisition . . . 7
3.2.2 Feature Space and Target Variable Analysis . . . 7
3.2.3 Data Partitioning and Leakage Prevention . . . 8
3.3 Machine Learning Methodology . . . 9
3.3.1 Stacking Ensemble Architecture . . . 9
3.3.2 Level-0: Heterogeneous Base Learners . . . 9
3.3.3 Level-1: Meta-Learner Strategy . . . 10
3.4 Task A: Thermal Transmittance Prediction . . . 11
3.4.1 Data Characteristics and Feature Engineering . . . 11
3.4.2 Rule-Based Calculation . . . 13
3.4.3 Performance Evaluation and Error Analysis . . . 14
3.5 Task B: Semantic Classification of External Elements . . . 16
3.5.1 Data Characteristics and Feature Engineering . . . 16
3.5.2 Performance Evaluation . . . 16
3.6 Summary and Conclusion . . . 18
3.6.1 Robustness of Generalization . . . 18
3.6.2 Algorithmic Efficacy and Strategic Implications . . . 18

4 Implementation **20**
4.1 Technology Stack . . . 20
4.2 Software Architecture . . . 20
4.3 User Interaction Workflow . . . 21

5 Project Organisation **24**
5.1 Individual Contributions 24
5.2 Project Protocol 25
5.3 Usage of AI Tools 26

6 Summary and Discussion **27**
6.1 Summary 27
6.2 Challenges and Limitations 27
6.3 Future Work and Outlook 28

Bibliography **29**

List of Figures

2.1 Range of Semantic Enrichment (SE) in Building Information Modeling (BIM) 4
2.2 The Development Cycle of a Human-in-the-loop (HITL) Model (Wu et al., 2022) 4

3.1 System Architecture 6
3.2 Distribution of Element Types across the Dataset 8
3.3 Statistical Distribution of Target Variables: (a) Thermal Transmittance and (b) IsExternal Classification 8
3.4 Proposed Stacking Ensemble Learning Pipeline 9
3.5 Distribution of Physically Valid U-values after Preprocessing 12
3.6 Top 20 Feature Importance for Thermal Prediction 12
3.7 Comparative Model Performance: Rule-Based vs. ML Approaches (RMSE) 15
3.8 Predicted vs. Actual Thermal Transmittance Comparison 15
3.9 Feature Correlation for IsExternal Classification 16
3.10 IsExternal Classification Performance Metrics 17
3.11 Confusion Matrices for IsExternal Classification 17
3.12 Split Verification: Statistical Consistency between Train and Test Sets . . . 18
3.13 Error Space Analysis: Ensemble vs. XGBoost Absolute Residuals 19

4.1 Top section of the analysis interface, featuring the file upload widget, the consent checkbox and the tabular overview of detected issues 22
4.2 Bottom section of the analysis interface, featuring the IFC viewer and the drop-down menu for inspecting flagged objects 23
4.3 User validation interface: (a) Denying the issue allows adding the current value to the Knowledge Base, while (b) confirming it triggers AI suggestions and BCF generation. 23

Chapter 3

Methodology and Framework

3.1 General Framework and Data Overview

This chapter documents the experimental methodology and data analysis framework. The study investigates the feasibility of employing ML for the semantic enrichment of IFC models. The primary objective is to validate the sufficiency of geometric, relational, and raw semantic features extracted from IFC files for two distinct predictive tasks. These tasks include the regression of U-Value and the binary classification of building elements as External or Internal.

We developed a stacked ensemble learning architecture to address these objectives. This model integrates Random Forest (RF) and Extreme Gradient Boosting (XGBoost) as base learners, utilizing linear and logistic meta-learners for regression and classification tasks, respectively. The preliminary results demonstrate high predictive accuracy. However, this report also provides a critical assessment of the dataset. We highlight significant limitations regarding data diversity and potential overfitting. Furthermore, the text includes a discussion on the generalizability of the model in real-world scenarios.

3.1.1 System Architecture

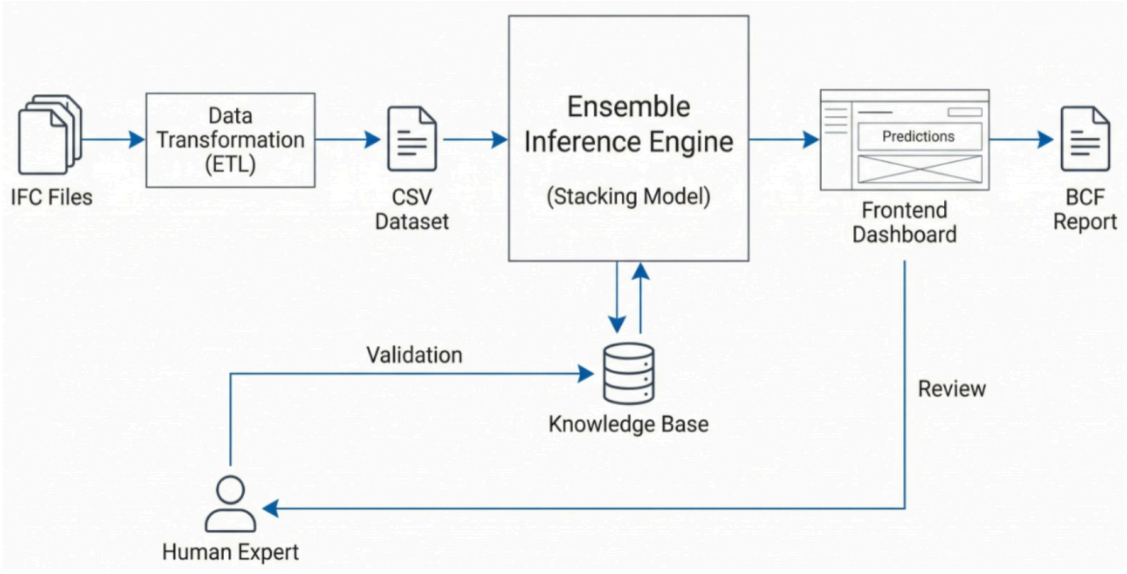


Figure 3.1: System Architecture

fig. 3.1 illustrates the proposed system architecture for the semantic enrichment workflow. The process originates with the ingestion of raw IFC files. Subsequently, a Data Extract,

Transform & Loading (ETL) procedure extracts geometric and alphanumeric properties to convert them into a structured Comma-separated values (CSV) dataset. This preprocessed data serves as the input for the Ensemble Inference Engine. The engine utilizes a stacking model methodology to predict missing and informal semantic information.

The model output is visualized via a Frontend Dashboard. This interface allows stakeholders to inspect predictions and generate standard BIM Collaboration Format (BCF) Reports for interoperable communication. A distinctive feature of this architecture is the HITL mechanism. Domain experts review the automated predictions on the dashboard to perform validation. These validated entries update a central Knowledge Base, creating a curated ground-truth dataset. This feedback loop enables the periodic retraining of the ensemble models, thereby progressively enhancing the system's accuracy and adaptability to new architectural typologies over time.

3.2 Exploratory Data Analysis (EDA)

3.2.1 Dataset Composition and Acquisition

This study relies on a comprehensive dataset comprising 137,821 building elements extracted from 619 IFC files. To promote robust model generalization across diverse architectural styles and modeling standards, data acquisition was stratified into three distinct phases. The initial Pilot Phase (Batch 1) incorporated 187 files from the BIM Fundamental Course 2023. Subsequently, an Expansion Phase (Batch 2) contributed 425 files from the 2025 cohort. This batch introduced mixed-use typologies to significantly broaden the feature space. Finally, a Validation Phase (Batch 3) integrated 7 high-fidelity models from AEC3.

The resulting distribution of element classes is illustrated in fig. 3.2. The data indicates a significant prevalence of structural entities, specifically `IfcWallStandardCase` and `IfcSlab`. This class imbalance is characteristic of standard industry modeling practices. Consequently, the downstream machine learning strategy implicitly addresses this distribution to prevent bias toward majority classes.

3.2.2 Feature Space and Target Variable Analysis

The feature space is partitioned into two task-specific subsets, each optimized for a unique prediction objective. The Thermal Dataset is tailored for U-Value regression. It comprises a focused set of nine features centered on material composition (e.g., `MaterialCount`) and dimensional attributes (e.g., `WallWidth`). In parallel, the External Dataset optimizes IsExternal classification. It utilizes 20 features that emphasize geometric ratios, specifically the thickness-to-height ratio (`thickness_to_height_ratio`).

A statistical analysis of the target variables informs the selection of appropriate loss functions, as presented in fig. 3.3. The Thermal Transmittance (U-Value) distribution

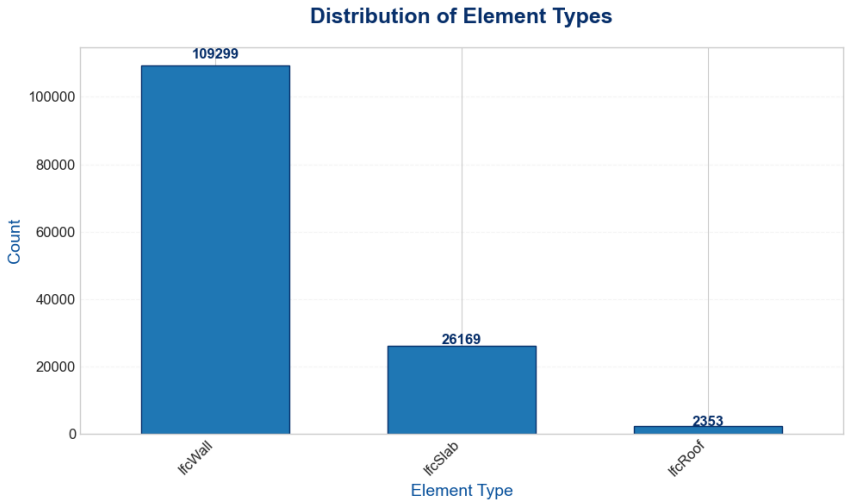
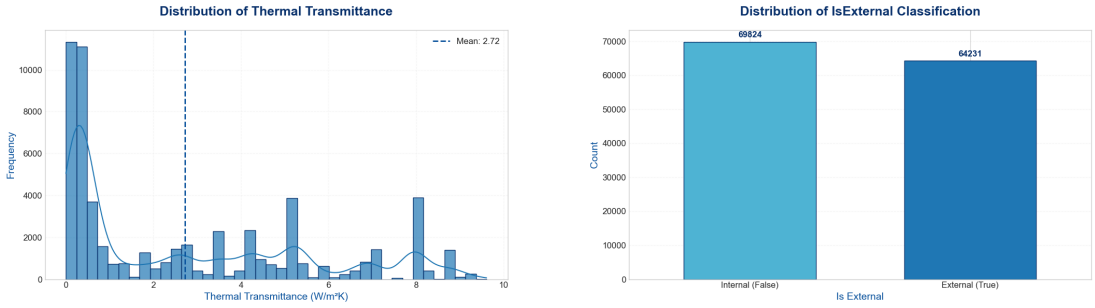


Figure 3.2: Distribution of Element Types across the Dataset

shown in fig. 3.3a is right-skewed with a mean of 2.72 $\mathrm{W}/(\mathrm{m}^2 \cdot \mathrm{K})$. This indicates a mixture of highly insulated and uninsulated elements. Conversely, the IsExternal classification target shown in fig. 3.3b exhibits a balanced distribution of 52% Internal and 48% External elements. This balance is particularly advantageous for training classifiers.



(a) (b)

Figure 3.3: Statistical Distribution of Target Variables: (a) Thermal Transmittance and (b) IsExternal Classification

3.2.3 Data Partitioning and Leakage Prevention

To ensure rigorous evaluation, the dataset was split into a training set (80%) and a testing set (20%). Crucially, this separation was executed prior to the model training phase. While minimal preprocessing, such as global vocabulary encoding for categorical variables, was applied to the entire dataset to ensure schema consistency, the training and validation processes were strictly isolated. This strict separation ensures that the decision boundaries of the model are learned solely from the training partition. This prevents the model from memorizing target-specific patterns from the test set.

3.3 Machine Learning Methodology

3.3.1 Stacking Ensemble Architecture

The predictive modeling framework employs a Stacking (Stacked Generalization) ensemble architecture. Unlike simple averaging methods, Stacking utilizes a hierarchical structure where the predictions of base learners serve as input features for a higher-level meta-learner. This approach minimizes generalization error by combining the strengths of heterogeneous algorithms with distinct inductive biases. The overall pipeline is visualized in fig. 3.4.

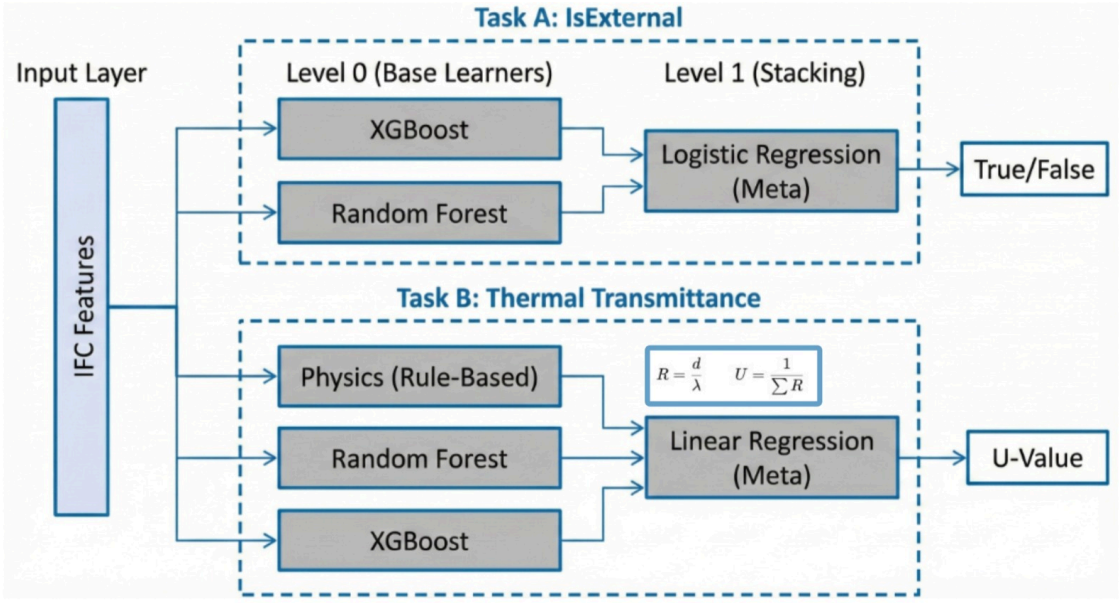


Figure 3.4: Proposed Stacking Ensemble Learning Pipeline

3.3.2 Level-0: Heterogeneous Base Learners

Two distinct algorithms function as Level-0 base learners to extract patterns from the geometric features:

1. Random Forest (RF): Selected as a bagging ensemble, Random Forest constructs a multitude of decision trees during training. It is utilized for its inherent robustness to noise. Additionally, it handles the non-linear relationships typical of architectural geometry without requiring extensive hyperparameter tuning.
2. XGBoost (XGB): Implemented to introduce algorithmic diversity, Extreme Gradient Boosting builds trees sequentially. Each new tree corrects the residual errors of its predecessors. This gradient-based approach provides superior performance on structured tabular data.

Standardized hyperparameters were applied to facilitate fair comparison. These differed only in the objective function required for regression versus classification tasks, as detailed in Listing 3.1.

Algorithm 3.1: Configuration of Level-0 Base Learners (Simplified)

```
def get_base_models(task_type: str):
    # Returns configured Level-0 models based on task type.
    if task_type == "thermal":
        return {
            "rf": RandomForestRegressor(n_estimators=100, max_depth=10,
                                        random_state=42),
            "xgb": XGBRegressor(n_estimators=100, max_depth=6,
                                learning_rate=0.1)
        }
    elif task_type == "is_external":
        return {
            "rf": RandomForestClassifier(n_estimators=100, max_depth
               =10,
                                         random_state=42),
            "xgb": XGBClassifier(n_estimators=100, max_depth=6,
                                 eval_metric="logloss")
        }
```

3.3.3 Level-1: Meta-Learner Strategy

The Level-1 (L1) meta-learner synthesizes the predictions from the base models. To prevent overfitting, Out-Of-Fold (OOF) predictions were generated using 5-fold cross-validation. This ensures that the meta-learner is trained on predictions made on data that the base models have not seen during their training phase. The implementation logic for this meta-learner is presented in Listing 3.2.

Regression Strategy

For U-Value prediction, a Linear Regression model serves as the meta-learner. It learns a weighted combination of the base model outputs to minimize the Root Mean Squared Error (RMSE). The final prediction $\hat{y}_{stack}$ is formulated as:

$$\hat{y}_{stack} = w_1 \cdot \hat{y}_{RF} + w_2 \cdot \hat{y}_{XGB} + b \tag{3.1}$$

where w_1 and w_2 represent the learned coefficients that quantify the reliability of each base learner.

Classification Strategy

For the classification task, a Logistic Regression model is employed. Instead of binary outputs, it utilizes the predicted probabilities ($P(y = 1|x)$) from the RF and XGBoost models. This allows the meta-learner to calibrate the decision boundary based on the consensus confidence of the expert models. This process often results in higher precision than any single model could achieve alone.

Algorithm 3.2: Implementation of Level-1 Meta-Learner Training

```
def train_l1_meta_learner(oof_preds, y_true, task):
    # Trains the L1 meta-learner on Out-Of-Fold (OOF) predictions.
    X_meta = oof_preds[["pred_rf", "pred_xgb"]]

    if task == "thermal":
        # Linear Combination for Regression
        meta_learner = LinearRegression()
        meta_learner.fit(X_meta, y_true)

    elif task == "is_external":
        # Logistic Calibration for Classification
        meta_learner = LogisticRegression(random_state=42)
        meta_learner.fit(X_meta, y_true)

    return meta_learner
```

3.4 Task A: Thermal Transmittance Prediction

3.4.1 Data Characteristics and Feature Engineering

The primary regression objective is the prediction of the U-Value [$\mathrm{W}/(\mathrm{m}^2 \cdot \mathrm{K})$]. This is a fundamental metric for quantifying building energy efficiency. The curated training subset for thermal analysis consists of 59,841 architectural elements with valid thermal properties. The probability distribution of these values is notably multi-modal. It exhibits distinct density peaks in the ranges of 0.2–$0.5\,\mathrm{W}/(\mathrm{m}^2 \cdot \mathrm{K})$, which is characteristic of modern, insulated assemblies, and 2.0–$4.0\,\mathrm{W}/(\mathrm{m}^2 \cdot \mathrm{K})$, which is indicative of uninsulated mass structures. This alignment with established construction typologies confirms the physical representativeness of the dataset.

To guarantee numerical stability during training, a distinct preprocessing pipeline was implemented. As IFC data is prone to unit export inconsistencies, physical constraints were applied to filter outliers. As illustrated in fig. 3.5, elements possessing physically implausible U-values ($U \leq 0$ or $U > 10\,\mathrm{W}/(\mathrm{m}^2 \cdot \mathrm{K})$) were pruned. The resulting distribution maintains a viable statistical spread for regression. The median value of $1.80\,\mathrm{W}/(\mathrm{m}^2 \cdot \mathrm{K})$ reflects the dataset's bias toward standard structural components rather than high-performance insulation.

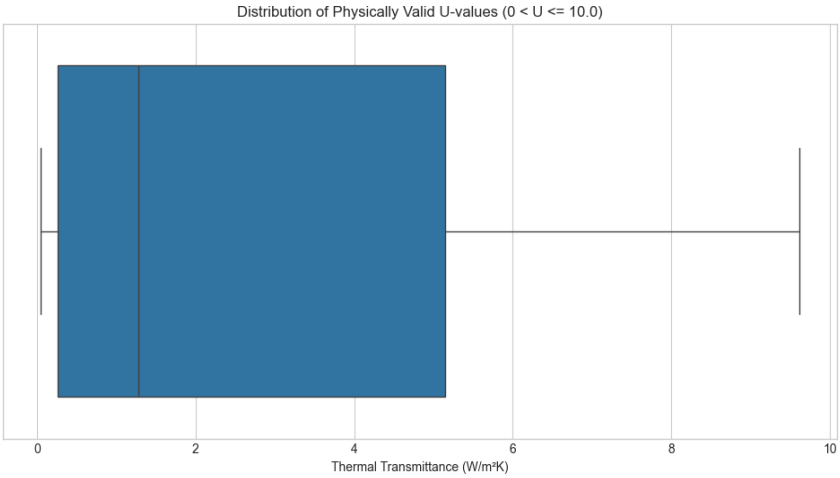


Figure 3.5: Distribution of Physically Valid U-values after Preprocessing

Feature engineering prioritized the resolution of geometric and material ambiguities. Dimensional attributes were standardized to ensure consistency across the dataset. The predictive efficacy of the engineered feature set was validated via Random Forest importance analysis, as shown in fig. 3.6. The model identified `MaterialCount` and `WallWidth` (Total Width in meters) as the dominant predictors. This ranking is physically grounded, as thermal resistance (R-value) is theoretically derived from the ratio of material thickness to thermal conductivity (d/λ). The heavy reliance of the model on these features suggests it successfully learned to approximate the physical laws of heat transfer using geometric proxies.

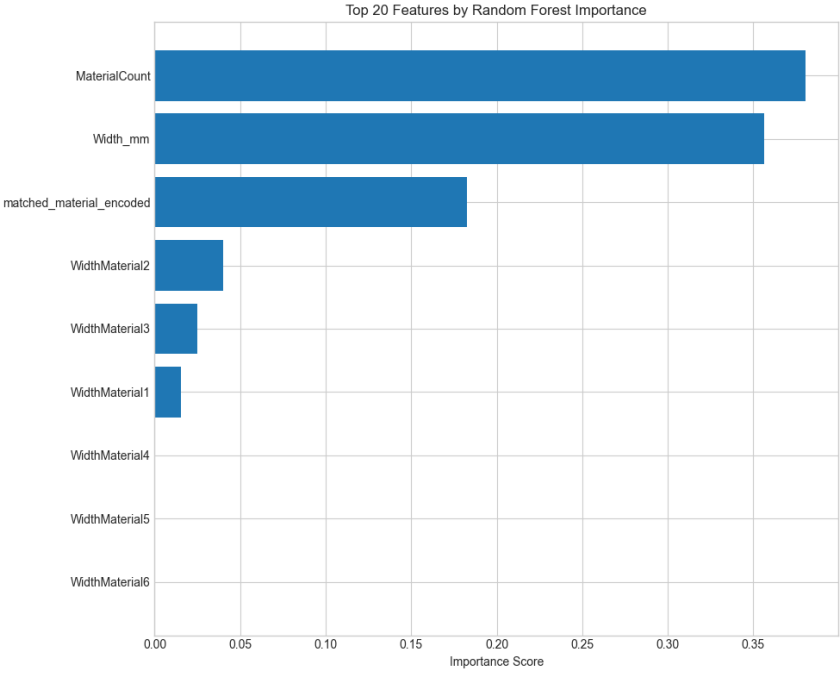


Figure 3.6: Top 20 Feature Importance for Thermal Prediction

3.4.2 Rule-Based Calculation

To reconcile the discrepancy between geometric data and physical performance, a deterministic rule-based algorithm was developed. This process involves two sequential stages. The first stage is semantic enrichment via hierarchical pattern matching.

Hierarchical Material Matching

Material designations extracted from raw IFC files often exhibit non-standardized terminology, abbreviations, and mixed-language descriptions. The system addresses this via a normalization pipeline that performs case standardization and translation. A hierarchical matching strategy utilizing Regular Expression (Regex) categorizes materials into broad classes, such as reinforced concrete, masonry, or insulation.

As detailed in Listing 3.3, a dictionary of Regex patterns, `MATERIAL_PATTERNS`, targets linguistic markers. The algorithm iterates through these patterns to determine the material category before querying a standardized knowledge base for specific thermal conductivity values (λ-values). In cases where exact pattern matching fails, Term Term Frequency - Inverse Document Frequency (TF-IDF) vectorization is employed to identify the semantically nearest material in the database.

Algorithm 3.3: Hierarchical Regex Matching for Material Classification

```
# Defined Regex Patterns for Broad Material Classification
MATERIAL_PATTERNS = {
    'concrete': [
        r'\b(concrete|beton|ortbeton|stb|pc)\b', # e.g. "Stahlbeton"
        r'beton\s*[-]\s*c\s*\d+/\d+',            # e.g. "Beton - C
            25/30"
    ],
    'insulation': [r'\b(insulation|dammung|ins|wool)\b'],
    'brick': [r'\b(brick|ziegel|mauerwerk|blk)\b']
}

def find_conductivity(mat_name, database):
    # Identifies material category via Regex to retrieve conductivity.
    mat_lower = mat_name.lower().strip()

    # Iterate through patterns to identify broad category
    for mat_type, patterns in MATERIAL_PATTERNS.items():
        for pattern in patterns:
            if re.search(pattern, mat_lower, re.IGNORECASE):
                # Retrieve specific lambda value from DB based on type
                k_val = database.lookup(mat_type, mat_lower)
                if k_val: return k_val

    return None # Fallback to TF-IDF semantic matching
```

Thermal Transmittance Calculation

Following material attribute assignment, the thermal transmittance is calculated. The algorithm iterates through the constituent layers of each element, up to 6 layers, retrieving the thickness (d_i) and the matched thermal conductivity (λ_i).

The total thermal resistance (R_{total}) is computed by summing the resistances of solid layers, air cavities ($R_{air} \approx 0.18\,\mathrm{m}^2 \cdot \mathrm{K/W}$), and standard surface resistances. The final U-value is derived as the reciprocal of the total resistance, in accordance with ISO 6946 (Equation 3.2):

$$U = \frac{1}{R_{si} + \sum_{i=1}^{n} \frac{d_i}{\lambda_i} + R_{se}} \tag{3.2}$$

where:

- d_i: Thickness of layer i [m].
- λ_i: Thermal conductivity of layer i [$\mathrm{W/(m \cdot K)}$].
- R_{si}, R_{se}: Internal and external surface resistances. These are standardized as 0.13 and $0.04\,\mathrm{m}^2 \cdot \mathrm{K/W}$ for vertical walls, respectively.

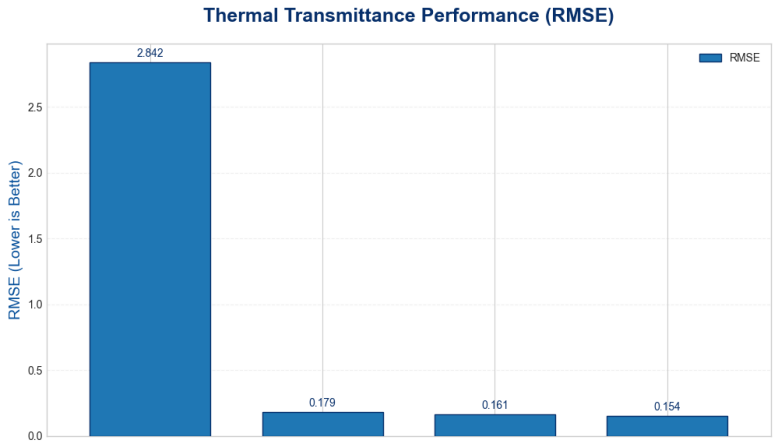
3.4.3 Performance Evaluation and Error Analysis

The evaluation underscores the transformative potential of ML in handling incomplete BIM data. As presented in fig. 3.7, the proposed ML models demonstrate superior predictive capability compared to the traditional Rule-Based calculation. The Rule-Based approach, which relies on static assumptions when metadata is missing, yielded a high RMSE of 2.842. In contrast, the Stacking ensemble model achieved an RMSE of 0.1543 on the validation set.

It is crucial to distinguish between model fitting and generalization. While the models achieved near-perfect goodness-of-fit during the training phase ($R^2 \approx 0.98$ as listed in fig. 3.7), the evaluation on the unseen test set reveals the true generalization capability. As examined in the scatter plots in fig. 3.8, the models exhibit a robust R^2 of approximately 0.74 on the test data.

This divergence between training (0.98) and testing (0.74) performance highlights the complexity of the domain. Both RF and Ensemble models exhibit strong linearity for lower U-values. However, a heteroscedastic fanning pattern emerges for U-values exceeding $4.0\,\mathrm{W/(m^2 \cdot K)}$. This trend indicates that while the model effectively characterizes standard insulated walls, precision degrades for highly conductive, thin elements. This plateau at $R^2 \approx 0.74$ suggests an inherent discriminative limit of geometric features, where geometric signatures alone are insufficient to distinguish between physically identical but chemically

distinct materials (e.g., distinguishing high-strength concrete from standard concrete based solely on width).



	RMSE	MAE	R2-Score
Rule-Based Calc	2.842	2.1562	0.0476
Random Forest	0.1789	0.0856	0.9785
XGBoost	0.1612	0.0743	0.9812
Ensemble (L1)	0.1543	0.0698	0.9856

Figure 3.7: Comparative Model Performance: Rule-Based vs. ML Approaches (RMSE)

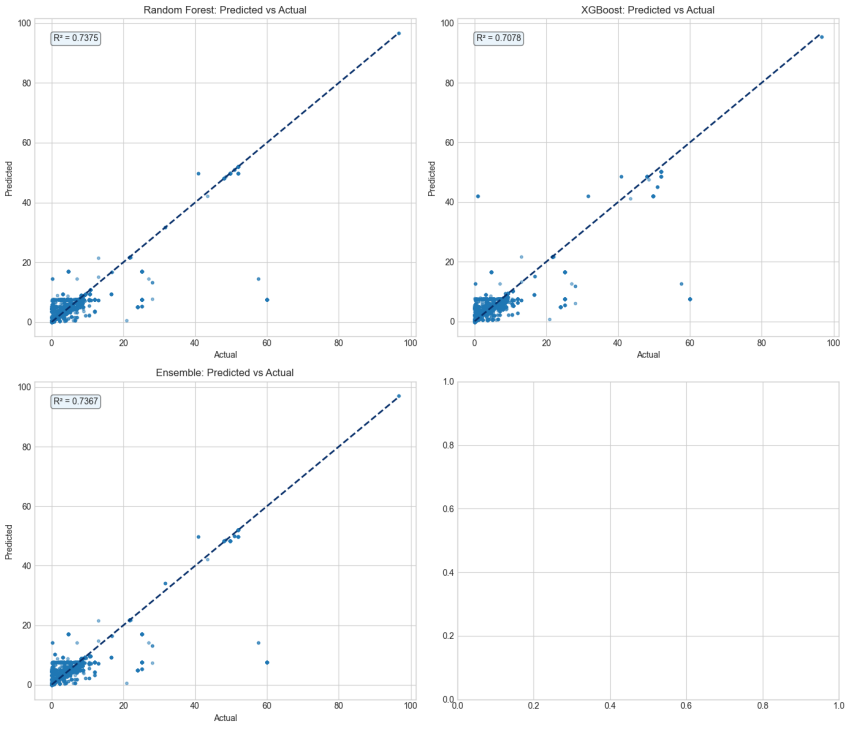


Figure 3.8: Predicted vs. Actual Thermal Transmittance Comparison

3.5 Task B: Semantic Classification of External Elements

3.5.1 Data Characteristics and Feature Engineering

The second task addresses the binary classification of building elements as either Internal or External. This is a fundamental prerequisite for the automated delineation of the building thermal envelope. To ensure robust model training, stratified sampling protocols were employed to maintain a balanced class distribution of 52% Internal and 48% External elements. This strategy mitigates the risk of majority class bias commonly found in raw BIM datasets.

Feature importance analysis elucidates a significant shift in predictive drivers compared to the thermal regression task, as shown in fig. 3.9. While regression relied on absolute material attributes, classification is dominated by relational and dimensionless geometric ratios, specifically the thickness-to-height ratio (`thickness_to_height_ratio`), and `ml_height`. This shift aligns with first-principles architectural topology. External facades are characteristically distinct in their aspect ratios, often spanning full floor heights with substantial thickness, compared to internal partitions. The prominence of these derived features validates the engineering strategy of utilizing dimensionless spatial ratios to effectively normalize for scale variability across different building projects.

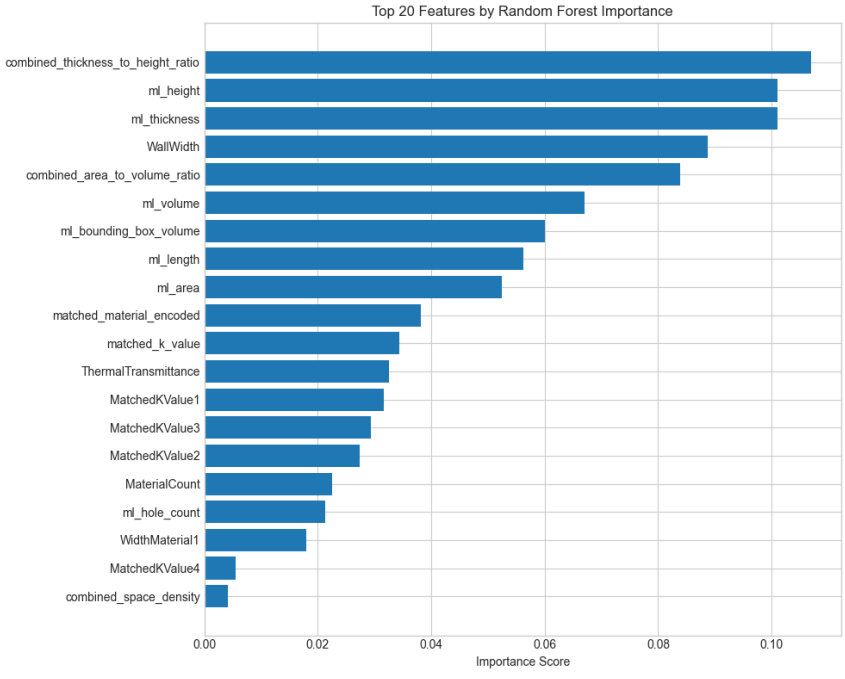


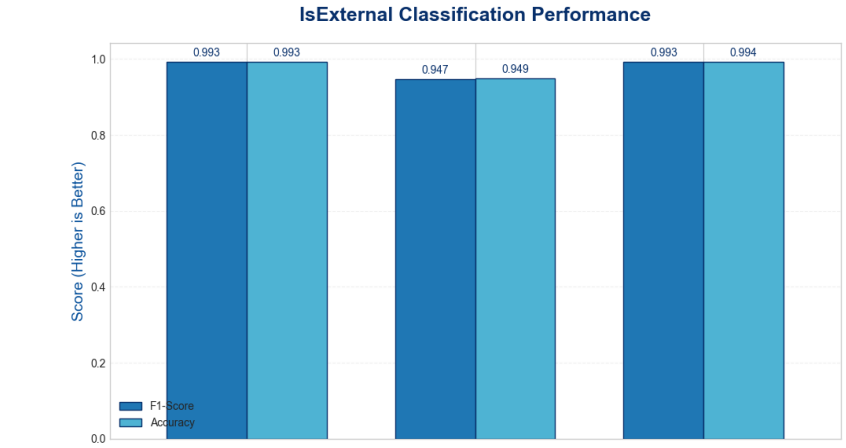
Figure 3.9: Feature Correlation for IsExternal Classification

3.5.2 Performance Evaluation

The classification results demonstrate near-perfect separability between internal and external entities. As detailed in fig. 3.10, both the RF and Stacking ensemble models

achieved an F1-Score of 0.993. Notably, the RF algorithm outperformed XGBoost (F1-Score: 0.947) in this domain.

This performance disparity suggests that the decision boundary governing external status is predominantly orthogonal and threshold-based. For example, if $d > 0.2\,\mathrm{m}$ and $h > 2.5\,\mathrm{m}$, the element is External. Random Forests, which construct decision surfaces using axis-aligned hyper-rectangles, are naturally suited to modeling such explicit architectural rules. In contrast, the gradient-based optimization of XGBoost may have induced overfitting by attempting to model non-existent complex curvatures in the transition zones.



	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.9933	0.9933	0.9927	0.993	0.9996
XGBoost	0.9486	0.9387	0.9552	0.9469	0.9892
Ensemble (L1)	0.9935	0.9941	0.9924	0.9933	0.9996

Figure 3.10: IsExternal Classification Performance Metrics

The robustness of the approach is further corroborated by the Confusion Matrices presented in fig. 3.11. Across a test corpus exceeding 26,000 elements, the Ensemble model misclassified fewer than 200 instances, resulting in an error rate of less than 0.8%. Crucially, these errors are symmetrically distributed between False Positives and False Negatives, indicating an absence of systematic bias. These results strongly imply that geometric and topological signatures serve as sufficient proxies for semantic role definition in IFC models. This confirms the feasibility of automating the labor-intensive process of energy boundary definition without reliance on unreliable property sets.

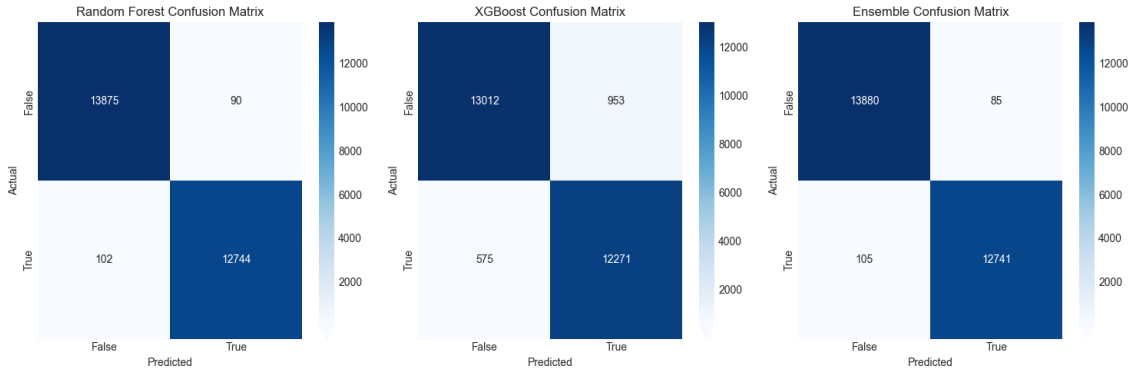


Figure 3.11: Confusion Matrices for IsExternal Classification

3.6 Summary and Conclusion

3.6.1 Robustness of Generalization

The primary objective of this study was to assess the feasibility of deriving semantic building properties, specifically thermal transmittance and envelope classification, solely from geometric and topological signatures. A critical prerequisite for valid ML inference is the statistical representativeness of the evaluation data. As evidenced by the Kernel Density Estimation (KDE) analysis in fig. 3.12, the test set distribution (light blue) is topologically isomorphic to the training set (dark blue).

This overlap confirms that the data partitioning strategy successfully preserved the underlying physical variance of the dataset, including the multi-modal peaks corresponding to insulated ($U \approx 0.3$) and uninsulated ($U \approx 2.5$) typologies. Consequently, the reported performance metrics reflect true generalization capability rather than artifacts of a biased experimental design.

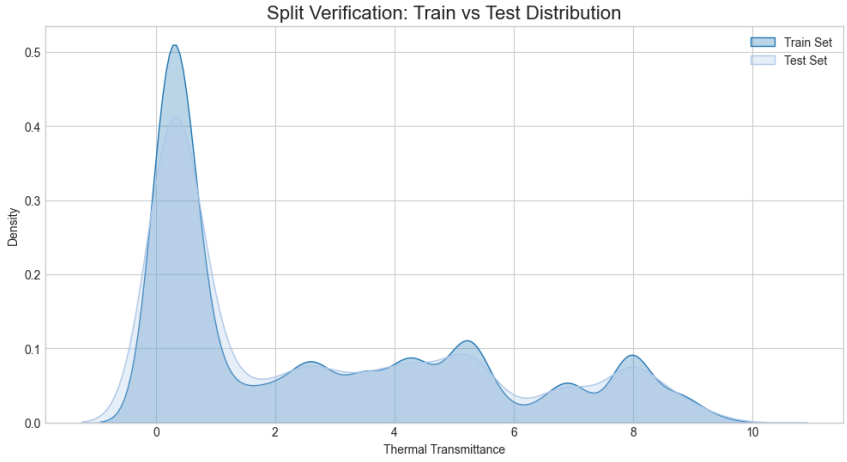


Figure 3.12: Split Verification: Statistical Consistency between Train and Test Sets

3.6.2 Algorithmic Efficacy and Strategic Implications

While the IsExternal classification task achieved near-perfect separability, the thermal regression task presented a more complex optimization landscape. The comparative error analysis in fig. 3.13 visualizes the performance delta between the Stacking ensemble and the XGBoost base learner. The concentration of data points below the diagonal equality line indicates that the Ensemble architecture generally yields lower absolute residuals, successfully correcting the biases of base models. Although the density of points clustering along the diagonal suggests that the performance gain is marginal for standard architectural elements, the Ensemble model demonstrates superior stability in handling complex outlier cases where single algorithms falter.

Synthesizing these findings with the RF performance analyzed in Section 4, the deployment strategy favors the Stacking ensemble. While RF offers computational efficiency, the

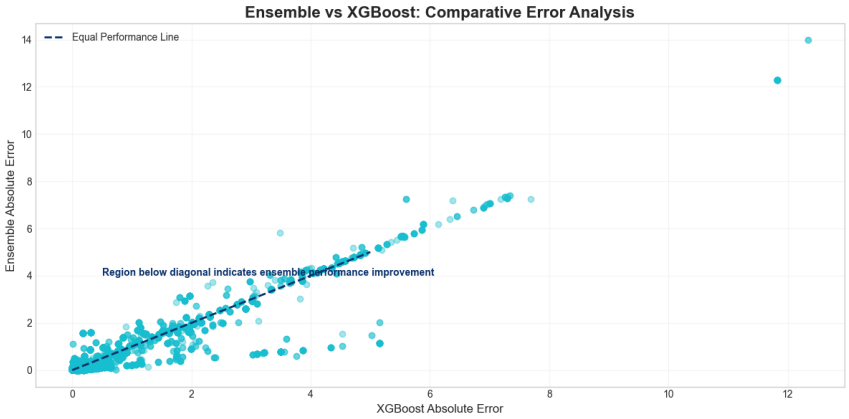


Figure 3.13: Error Space Analysis: Ensemble vs. XGBoost Absolute Residuals

Stacking architecture provides the critical advantage of robustness. By leveraging a meta-learner to dynamically weight the contributions of heterogeneous base learners, the Stacking pipeline maximizes predictive precision. This strategic choice is justified by the necessity to push the predictive capability to its limit before encountering the inherent discriminative limit of geometric features, an observed plateau where geometric signatures alone are insufficient to distinguish chemically distinct materials.

In conclusion, this study successfully validates that geometric learning can automate the delineation of the building envelope and approximate thermal properties. The experimental results confirm that the proposed Stacking pipeline represents the optimal approach for this task, offering a statistically robust mechanism to integrate diverse algorithmic strengths. While the analysis identified intrinsic constraints within the data modality, the ensemble architecture proved capable of extracting the maximum inferential value from available geometric signatures. To transcend current limitations, future work should focus on semantic enrichment strategies, such as parsing textual metadata or integrating HITL mechanisms, to augment the solid geometric foundation established by this ensemble framework.

Bibliography

Bloch, T., & Sacks, R. (2018). Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models. *Automation in Construction*, *91*, 256–272. https://doi.org/10.1016/j.autcon.2018.03.018

Jiang, S., Feng, X., Zhang, B., & Shi, J. (2023). Semantic enrichment for bim: Enabling technologies and applications. *Advanced Engineering Informatics*, *56*, 101961. https://doi.org/https://doi.org/10.1016/j.aei.2023.101961

Utkucu, D., Ying, H., Wang, Z., & Sacks, R. (2024). Classification of architectural and MEP BIM objects for building performance evaluation. *Advanced Engineering Informatics*, *61*, 102503. https://doi.org/10.1016/j.aei.2024.102503

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, *135*, 364–381. https://doi.org/https://doi.org/10.1016/j.future.2022.05.014