

Article

Not peer-reviewed version

Automatic Classification of BIM Object Based on IFC Data Using the Uniclass Classification Standard

[Shi Tang](#)*, [Takamasa Bito](#), [Kazuya Shide](#)

Posted Date: 23 May 2025

doi: 10.20944/preprints202505.1790.v1

Keywords: Building Information Modeling; IFC; Random Forest; Uniclass; Classification



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Automatic Classification of BIM Object Based on IFC Data Using the Uniclass Classification Standard

Shi Tang ^{1,*}, Takamasa Bito ² and Kazuya Shide ³

¹ Doctoral Course, Graduate School of Eng., Shibaura Institute of Technology

² STARTS Research Institute, Ltd.

³ School of Architecture, Shibaura Institute of Technology

* Correspondence: na24104@shibaura-it.ac.jp; Tel.: +81-070-4156-9289

Abstract: Classification of BIM objects is critical for enhancing information interoperability and standardization within construction projects; however, research on automated BIM object classification based on standardized classification systems remains limited. Therefore, this study aims to develop an automated classification method based on the Uniclass system, utilizing low-LOD IFC data combined with machine learning algorithms to achieve standardized classification of BIM objects. The proposed method first assigns Uniclass codes to BIM objects, then extracts IFC data and coding information as feature variables and classification labels, and finally applies a Random Forest model to perform automatic classification. Experimental results demonstrate that the proposed method achieves classification accuracies of 1.00 and 0.99 for BIM objects at the Elements/Functions and Systems coding tasks. This model enables the automatic classification of BIM objects relying solely on low-LOD input, thereby reducing dependence on high-detail modeling and improving applicability during the early design phase.

Keywords: building information modeling; IFC; random forest; uniclass; classification

1. Introduction

In the Architecture, Engineering, and Construction (AEC) industry, the implementation of building projects requires a high degree of collaboration and seamless information flow [1]. Building Information Modeling (BIM) serves as a digital representation of the information contained within a building. BIM models, characterized by their rich and well-structured data, facilitate efficient information collaboration throughout the entire lifecycle of a construction project, thereby improving design and construction efficiency while effectively reducing waste and cost overruns [2]. Within the BIM environment, the Industry Foundation Classes (IFC) format functions as a centralized data storage medium for building projects [3]. Compared to traditional construction workflows, IFC data can be integrated with cutting-edge technologies such as Big Data, Artificial Intelligence (AI), the Internet of Things (IoT), unmanned aerial vehicle (UAV) technologies, and Geographic Information Systems (GIS), thereby promoting the advanced application and development of BIM in construction engineering projects [4].

To support the application of such advanced technologies, BIM objects should be classified to enable structured information management, thereby improving their usability and maintainability [5]. As an international standard classification system in the architecture domain, Uniclass provides a coding framework and hierarchical structure for systematically organizing components, functions, and other elements [6]. This enhances the consistency and interoperability of information management and facilitates the fine-grained management and efficient reuse of information. Pupeikis et al. conducted a comparative study between Denmark's Construction Classification System (CCI) and the United Kingdom's Uniclass 2015 in the context of BIM applications. Their findings indicated that Uniclass 2015 offers certain advantages in terms of classification structure and applicability, enabling more effective information management and collaboration in BIM projects [7]. Although

some studies have attempted to encode prefabricated components, traditional coding methods are often time-consuming, labor-intensive, and prone to errors [8]. L. Choi et al. proposed a rule-based automatic classification method, which improves efficiency compared to manual classification [9]. However, rule-based classification heavily relies on manually defined rules or conditions, typically grounded in expert knowledge or pre-established logic. When data characteristics or task scenarios change, the rule sets must be manually updated. Moreover, the creation of rule libraries depends significantly on highly experienced experts, whose knowledge is subjective and difficult to standardize or share. In recent years, with the development of artificial intelligence, classification methods based on image recognition and semantic analysis have emerged, leading to significant improvements in the accuracy and adaptability of automatic classification approaches [10]. Liu et al. proposed an "Automatic Fine-Grained BIM Element Classification" method, which aims to automate the detailed classification of elements within BIM. The study employs a multimodal deep learning (MMDL) approach, integrating geometric and semantic information to enhance classification accuracy and efficiency. This method enables the automatic classification of BIM elements, thereby improving information management and decision-making support during the design and construction processes [11]. M. Seydgar evaluated the performance of various deep neural network models based on point clouds for IFC object classification, demonstrating excellent results in the automatic classification of BIM information [12]. However, these methods primarily focus on classifying the geometric shapes of construction components and do not adopt a standardized classification framework. Xu et al. developed an automated classification and coding system using IFC data as feature variables within BIM models, achieving unique identification and information sharing for prefabricated components [13]. While this study was the first to apply a standardized classification approach to BIM elements, its scope was limited to prefabricated components and was not applicable to all types of BIM objects. The comprehensive classification of BIM information based on standardized classification systems remains insufficiently explored in previous research.

Therefore, this study aims to develop a classification method based on the Uniclass classification system combined with machine learning algorithms to achieve the automatic classification of BIM objects. The experimental framework begins with the collection of BIM data, during which Uniclass codes are assigned to BIM objects. Subsequently, IFC information is extracted from BIM components to select and create the feature variables for the dataset. An appropriate machine learning classification model is then selected and trained using the dataset to classify BIM elements into fine-grained categories at Level 4 within the Uniclass hierarchy. Finally, the effectiveness of the Random Forest algorithm is validated by comparing its performance against other classification models.

The remainder of this paper is organized as follows: Section 2 reviews existing research on BIM object classification, classification methods based on building information classification standards, and the application of the Random Forest algorithm for data classification. Section 3 details the proposed methodological framework, including the Uniclass Elements/Functions (EF) and Systems (Ss) classification structures adopted in this study, the data collection and feature selection methods for machine learning model training, and the experimental analysis methods. Section 4 presents the data preprocessing procedures, the classification models employed, and the experimental results. Section 5 discusses the experimental outcomes, summarizes the research contributions, compares the findings with existing studies, and outlines the study's limitations and future research directions. Finally, Section 6 concludes the paper.

2. Related Work

2.1. BIM Object Classification

The classification of BIM objects plays a crucial role in improving the organization and retrieval efficiency of construction information, facilitating information sharing and collaboration among project teams, reducing design and construction errors, and ultimately lowering costs and improving project quality [1,14]. Early research primarily relied on expert systems to address this issue. Zhi et

al. proposed a semi-explicit coding method for prefabricated building components, aiming to improve the dissemination and integration of information for prefabricated components in China through a flexible and widely applicable coding system [15]. Parece et al. developed a tool based on Building Information Modeling and a construction classification system to assess building carbon emissions. This tool integrated the classification system into BIM models, enabling adaptability across different development stages and modeling techniques, thereby simplifying the evaluation of design options and reducing the need for repeated modeling [16]. These studies focused on optimizing the organization and application of construction information through the integration of BIM technology and classification systems [15,16]. However, such methods have limited automation capabilities. When classification tasks involve a broad target range and complex data, it becomes difficult to ensure high efficiency and accuracy.

As a result, researchers have begun to explore data-driven methods, such as machine learning, which can learn and generalize classification rules from the data itself, thereby overcoming the efficiency and scalability limitations of rule-based approaches [8,17]. Utkucu et al. proposed an integrated method combining semantic keyword search, rule-based reasoning, machine learning based on object geometry, and deep learning based on visual shape features. This approach enabled automatic classification of architectural and MEP BIM objects, significantly improving the efficiency, interoperability, and accuracy of data exchange in building performance assessments [18]. Yu et al. proposed an ensemble deep learning method that combines a multi-view convolutional neural network (MVCNN) and a multilayer perceptron (MLP), leveraging image and adjacency features of BIM elements to enhance classification performance [19]. Koo et al. applied the support vector machine (SVM) algorithm to classify BIM objects automatically based on geometric and physical relationship features of construction elements, aiming to evaluate their semantic integrity [20]. These studies demonstrate that data-driven methods offer stronger interoperability, higher efficiency, and better classification accuracy in BIM object classification tasks [18–20]. However, challenges remain in terms of feature selection, model complexity, and data requirements. Most existing BIM object classification studies rely primarily on image-based datasets, which often lead to increased model complexity and a strong dependence on high-quality data.

2.2. Building Classification System

Architectural forms exhibit significant diversity in both form and function. Their basic components typically include structural systems, envelope elements, and interior finishes. In order to clarify the characteristics of buildings and ensure their appropriateness, it is essential to predefine construction-related information within a classification framework [21]. Building Information Classification Systems (BICS) began to gain widespread development in the United States and Europe from the early 20th century. Due to differences in national legislation, building regulations, and practical environments, the classification methods and structures adopted across regions have shown considerable diversity [2].

In 1948, Sweden developed the Sfb classification system, which incorporated practical building content and is considered one of the earliest architectural classification systems in the world. It was subsequently adopted by countries such as the United Kingdom. In 1963, the United States introduced the precursor to the MasterFormat system—CSI Format for Construction Specification. In 1987, the UK released the initial version of the Common Arrangement of Work Sections (CAWS). Additionally, the Royal Institute of British Architects (RIBA) refined and extended the Sfb system into the Sfb/UDC, and in 1968 proposed the Construction Index (CI)/Sfb system. Following these developments, the international standardization of construction classification systems was led by the International Organization for Standardization (ISO). In 1994, ISO published Technical Report ISO TR 14177, which addressed the classification of information in the building production process. Later, in 2001, ISO formally released ISO 12006-2: Building Construction — Organization of Information about Construction Works. To align with international standards, many countries subsequently restructured their national classification systems. For example, in 1997, the UK revised its CI/Sfb

system into Uniclass; in North America, MasterFormat and UniFormat were integrated to form OmniClass, based on ISO 12006-2. In Japan, the Japan Construction Information Center (JACIC) developed the JCCS (Japan Construction Classification System). Other countries also established systems tailored to their own construction contexts, such as StLB in Germany, CCS in Denmark, SfB+BSAB in Sweden, and Building 90 (Talo 90) in Finland. In summary, building classification systems across the globe have been gradually evolving in alignment with ISO 12006-2, forming a historically progressive trajectory of systematization. In recent years, with the continuous integration of information technologies into the construction industry, international standardization of classification systems has continued to evolve to meet the demands of digital transformation in building information. Especially with the widespread adoption of BIM technology, building classification systems have received increasing attention and are becoming essential tools for bridging project information with the building lifecycle.

In the United Kingdom, early practices traditionally relied on the CI/SfB Construction Information Index Handbook published by the RIBA in 1976, which was a faceted classification system composed of five classification tables. Subsequently, in 1987, RIBA, together with the RICS and other organizations, jointly established the Construction Project Information Committee (CPIc) and developed the Uniclass classification system in 1997. The system later underwent several major revisions: it evolved into Uniclass2 in 2013, was updated to Uniclass2015 in 2015, and most recently was renamed simply as Uniclass again in 2023, marking a total of four significant updates. Uniclass is applicable across all building-related fields, encompassing a wide range of projects from infrastructure such as railways to facility components like station surveillance cameras, all within a unified classification framework. With the introduction of BIM technologies, Uniclass provides a comprehensive and widely applicable classification system for infrastructure, landscape, engineering services, and building sectors, supporting all stages of the project lifecycle. Although both Uniclass and Omniclass are established classification systems developed in alignment with ISO 12006-2, this study adopts Uniclass 2015 due to its clearly defined hierarchical structure. As the subsequent stages of our research aim to conduct multi-level decomposition and analysis of building objects, Uniclass offers a more suitable framework for supporting structured classification and progressive refinement across different levels of BIM data.

As shown in Table 1, Uniclass consists of 15 classification tables organized hierarchically, comprehensively covering construction industry information from the most macro to the most micro levels.

Table 1. Uniclass Classification Tables is a table.

Uniclass Tables	Description
Activities (Ac)	The Activities table classifies the activities that take place in existing assets, or that need to be accommodated within them.
Complexes (Co)	The Complexes table classifies high-level groupings in the built environment and tends to describe a group of entities brought together in one place as a complex, for a particular purpose or multiple activities.
Entities (En)	The Entities table classifies individual parts of an asset, like buildings, bridges, or tunnels.
Spaces/ locations (SL)	The Spaces/ locations table classifies spaces where activities take place and locations where specific items or equipment can be found, often in linear infrastructure like pipelines, roads, and rail.
Elements/ functions (EF)	The Elements/ functions table classifies general elements like walls, decks, and roofs, which can be thought of as the main components of buildings, structures, towers or tunnels, and functions , which

	describe generic services required for asset operation such as piped gas supply, rail and paving heating, or waste collection.
Systems (Ss)	The Systems table classifies collections of products brought together to operate as systems , in order to provide a common purpose or solution.
Products (Pr)	The Products table classifies individual products used across the built environment, including those assembled to create systems, and objects located as part of asset operation or functions.
Tools and equipment (TE)	The Tools and equipment table classifies tools and equipment , such as plant machinery, vehicles, tunnel boring machines, formwork, scaffolding, and temporary hoardings across the full-range of built environment for the construction and ongoing maintenance and repair of assets.
Project management (PM)	The Project management table classifies requirements, information, and records for asset management and project management across the full lifecycle of the built environment, at all scales.
Form of information (FI)	The Form of information table classifies forms of information , often exchanged as part of asset management and construction projects, with codes for contract, quotation, room data sheet, bill of quantities, three-dimensional model, or invoice.
Roles (Ro)	The Roles table classifies the individual or organizational roles required in asset management and the successful delivery of built environment projects.
Risk (RK)	The Risk table is used to categorize various types of risks associated with the lifecycle of built assets, facilitating the identification, management, and communication of potential risks during the design, construction, and operation phases of a project.
Material (Ma)	The Material table classifies materials used in the built environment.
Properties and characteristics (PC)	The Properties and Characteristics table is designed to categorize various attributes and characteristics of built assets, supporting detailed description and management throughout the asset lifecycle, and enhancing consistency and traceability of information.
CAD and modelling content (Zz)	The Zz_ table supports CAD and modelling content to assist with clear and consistent layer naming in modelling platforms, and managing the various components required in digital drawings, models, and construction outputs.

Elements/Functions are used to classify components and functional elements within a building. "Elements" refer to the physical parts of a structure, such as floors, walls, and roofs, or major components of constructions like foundations and piers. "Functions" represent the services provided or managed by spaces or equipment. "Systems" are defined as assemblies of multiple components working together to form elements or deliver functions. For instance, in the case of a pitched roof, rafters, underlayments, tiles, ceiling boards, insulation materials, and ceiling finishes collectively constitute a system. Similarly, a low-temperature hot water heating system consists of boilers, pipes, water tanks, and radiators. With its hierarchical structure and comprehensive coverage, Uniclass enables the fine-grained organization of building-related information and meets the needs of BIM applications across all project stages. Therefore, this study adopts Uniclass as the classification

framework, primarily due to its clear structure, rigorous logic, and its ability to systematically support extensions and applications across different classification dimensions (such as EF, Ss, and the future integration of Pr).

2.3. Automatic Classification by Machine Learning

In recent years, with advancements in computational power and the availability of large datasets, machine learning has achieved significant breakthroughs across various fields [22–24]. Machine learning classification algorithms, through techniques such as natural language processing and classification modeling, not only enhance the accuracy of key variable quantification but also improve variable selection, topic extraction, and pattern recognition capabilities, thereby deepening the understanding of strategic management objects. In particular, classification models such as Random Forest and Support Vector Machine (SVM) have demonstrated substantial value in improving the effectiveness and robustness of model estimation [25]. Current research mainly focuses on numerical data, text classification, and image recognition tasks. In the field of construction engineering, Kim et al. applied a machine learning method based on terrestrial laser scanning (TLS) and density-driven techniques to automatically classify rebar diameters, thereby improving the automation and accuracy of rebar size detection [26]. Zeng et al. utilized machine learning methods to classify and predict the failure modes, load-bearing capacities, and effective stiffness of corroded reinforced concrete columns, enhancing the precision and efficiency of seismic performance assessments in construction engineering [27]. Li et al. employed machine learning-based classification methods to assess the defect conditions of wooden columns in ancient building walls, thereby achieving the automation of structural health monitoring in heritage structures [28]. Although machine learning-based classification methods have been applied to structural inspection, performance evaluation, and facility maintenance in the construction industry, the application of these techniques for the automatic classification of BIM components remains a crucial research topic.

Emunds et al. proposed SpaRSE-BIM, a sparse convolution-based neural network model designed to classify BIM geometric data in IFC format, enhancing the semantic richness of BIM models [29]. Koo et al. applied the SVM algorithm to automatically classify BIM model objects based on geometric and physical relationship features of building elements to verify semantic integrity [20]. Liu et al. developed a multimodal deep learning approach that integrates the geometric, semantic, and topological features of BIM objects to achieve fine-grained classification of building components [11]. These studies have demonstrated that machine learning methods can achieve high classification accuracy for BIM objects, enabling precise recognition and management of building information. However, research focusing on the classification of BIM objects based on standardized industry classification systems remains limited and warrants further exploration. Xu et al. introduced an automated classification and coding method based on BIM information and the Random Forest algorithm for prefabricated building components. Their method extracted IFC information from components and classified BIM objects according to the OmniClass classification system [13]. However, the feature variables required in their study included not only geometric and spatial information but also material attributes, which are often unavailable during the early design stages. Consequently, their method could not be implemented until the construction documentation phase. Additionally, their research focused solely on prefabricated components and did not extend to all types of building elements. Therefore, this study proposes an automatic classification method relying solely on low-detail IFC data as feature variables, aiming to achieve standardized classification of BIM objects based on the Uniclass classification system without depending on high-detail information or image data. This approach is intended to enhance the applicability and efficiency of classification methods during the early design stages of construction projects.

3. Methodology

3.1. Research Process

The research process is shown in Figure 1. First, to create the target labels for supervised learning, this study established a mapping relationship between BIM objects and the Uniclass classification system. The Uniclass EF and Ss codes were adopted as the standard classification basis, and a unique classification label was assigned to each IfcEntity, thereby generating the target variables required for model training. This process ensured the consistency and standardization of BIM object classification. Second, IFC4-format files were extracted from a set of BIM models to create the initial dataset. Using the BIM visualization tool BIMvision, the IFC files were parsed to extract IfcEntities representing individual building components. Each entity contained three categories of feature attributes: Semantic attributes, such as LoadBearing (whether load-bearing) and Storey (the associated floor); Spatial features, including global coordinates (GlobalX, GlobalY, GlobalZ) and bounding box dimensions; Dimensional information, such as maximum projected area, total area, and volume.

To address missing values in the original data, the K-Nearest Neighbors (KNN) interpolation method was applied for data imputation, resulting in a cleaned dataset. Subsequently, the dataset was randomly divided into training and testing subsets. The Random Forest algorithm was selected as the classification model, using the aforementioned input features to learn and predict the classification of BIM components, mapping them to their corresponding Uniclass codes. The model was trained in a batch-processing manner and evaluated on the testing set for classification performance. The final output established a mapping between BIM elements and their corresponding Uniclass codes — for example, classifying an IfcWall as Ef_25_10_40 (Internal walls) and Ss_25_10_30_35 (Gypsum board partition systems). Details of the methodology are provided in the following sections.

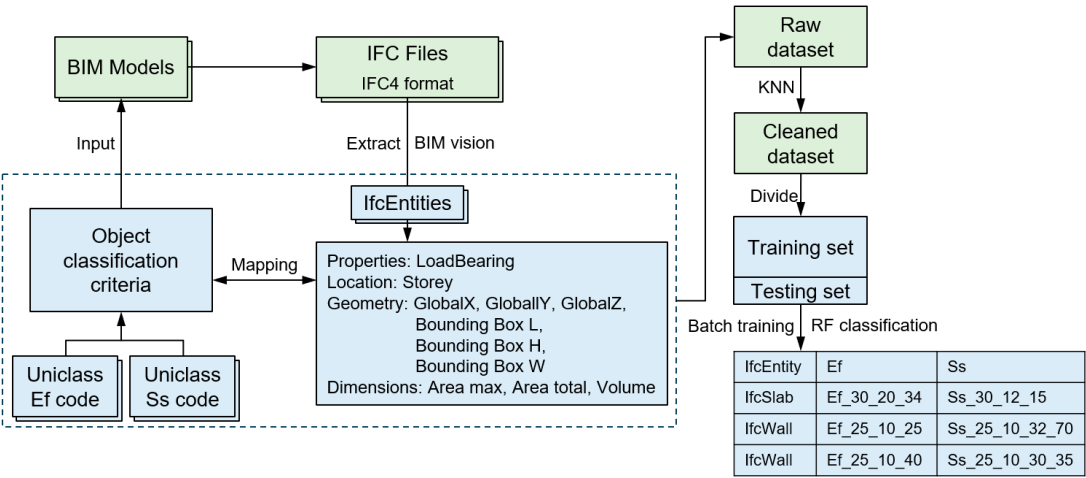


Figure 1. Overview of the proposed methodology framework.

3.2. Uniclass Classification System

In Uniclass, there exists a faceted structural relationship between the EF and SS tables, characterized by "A Part of" and "A Type of" associations, thereby achieving a unified hierarchical structure [30]. Specifically, the EF and SS tables share identical numbering at the level 1 (Group) and level 2 (Subgroup) hierarchies and establish a mapping relationship through "A Part of" associations. For example, as shown in Figure 2, for wall-related elements, the first-level entry in EF is "EF_25 Wall and barrier elements," corresponding to "Ss_25 Wall and barrier systems" in Ss; at the second level, "EF_25_10 Walls" includes several wall systems, such as "Ss_25_10 Framed wall systems" and "Ss_25_11 Monolithic wall structure systems." Thus, based on "EF_25_10 Walls," it is possible to further nest information such as "Ss_25_10_32 Framed wall systems," thereby establishing a hierarchical link between EF Level 2 and Ss Level 3. Moreover, EF Level 3 comprises various types such as External walls, Internal walls, and Parapet walls, forming a "Type of" structure; similarly, Ss

Level 4 presents a corresponding detailed structure. All EF and Ss codes involved in the dataset used in this study are listed in Table A1 in the Appendix. From the perspective of Level of Development (LOD)—a standard used to measure the granularity of component information—EF Levels 1 and 2 are more suitable for lower LOD stages (e.g., LOD 100–200), representing the basic functional classification of components. In contrast, EF Level 3 and Ss Level 4 are appropriate for higher LOD stages (e.g., LOD 300 and above), providing more detailed categorization of components and system divisions.

Therefore, from the viewpoint of decomposing a building into elemental components, a direct mapping between EF Level 2 and Ss Level 2 is reasonable. However, from an information management perspective based on breakdown structures (BS), it is necessary to insert the more detailed type information represented by EF Level 3 between EF Level 2 and Ss Level 2 to enhance the precision of information expression. Based on the above relationships, the hierarchical mapping logic between EF levels and Ss levels can be further analyzed, as illustrated in Figure 3.

Code	Description	Level	
Elements/Functions			
EF_25	Wall and barrier elements	Group	1
EF_25_10	Walls	Sub-group	2
EF_25_10_25	External walls	Section	3
EF_25_10_40	Internal walls	Section	3
EF_25_10_60	Parapet walls	Section	3
Systems			
Ss_25	Wall and barrier systems	Group	1
Ss_25_10	Framed wall systems	Sub-group	2
Ss_25_10_32	Framed wall structure systems	Section	3
Ss_25_10_32_03	Aluminium wall framing systems	Objects	4
Ss_25_10_32_65	Precast concrete wall systems	Objects	4
Ss_25_10_32_70	Reinforced concrete wall structure systems	Objects	4

Figure 2. Example of the Hierarchical Structure between EF and Ss Codes.

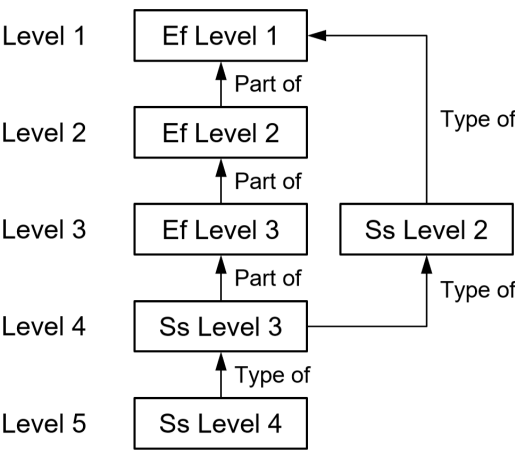


Figure 3. Hierarchical Relationship between EF and Ss Codes.

3.3. Data Acquisition and Feature Selection

The dataset used in this study was collected from five construction projects, all of which are residential buildings located in Japan. By gathering the BIM models of these projects and extracting the IFC data using the BIMvision software, each IfcEntity was treated as an individual data unit. In total, 8,715 data records were obtained, forming the dataset used for machine learning training. The basic information of these five projects is summarized in Table 2.

Table 2. Basic Information of the Construction Projects Used for Dataset Creation.

Project	Site Area(m ²)	Building Area(m ²)	Number of Floors	Structure Type	Structural Systems
A	221.44	167.00	6	RC (Seismic Resistant)	Rigid Frame
B ¹	-	-	-	-	-
C	463.65	261.73	9	RC (Seismic Isolation)	Rigid Frame
D	326.07	188.80	4	RC (Seismic Resistant)	Wall Structure
E	209.98	146.10	6	RC (Seismic Resistant)	Rigid Frame

¹ Project B consists of three complex buildings, and therefore its basic information is omitted.

To enable the automatic classification of BIM objects, this study created multiple feature variables from the extracted IFC data, primarily categorized into three types: semantic attributes, spatial features, and dimensional information. As illustrated in Figure 4 for a wall component, each IFC entity is not only assigned an EF code and an Ss code, but also described by the following features:

- Semantic attributes: including IfcEntity, Storey, LoadBearing, and IsExternal;
- Spatial features: including the global spatial coordinates within the model (GlobalX, GlobalY, GlobalZ) and the bounding box dimensions (length, width, and height);
- Dimensional information: including Area max, Area total, and Volume.

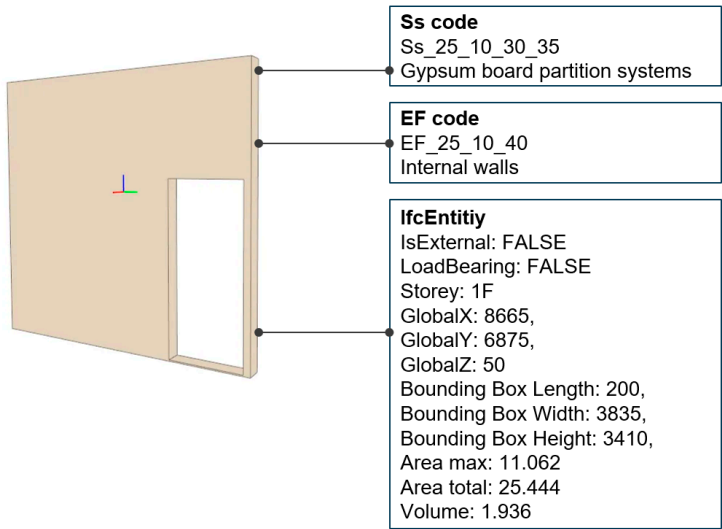


Figure 4. Feature Variables Extracted from IFC Data for Each BIM Object.

As shown in Figure 5, all input features correspond to information at LOD 200 or lower, which can be directly extracted from low-LOD IFC models without requiring highly detailed modeling information. In contrast, the inferred target variables (EF/Ss codes) correspond to classification information typically associated with approximately LOD 300. Therefore, this study establishes an automatic classification method that predicts higher-detail classification labels based on low-detail IFC data.

In addition, by utilizing the Autodesk Interoperability Tools plugin, shared parameters can be added to BIM objects within Revit models, thereby enabling the assignment of Uniclass-based EF and Ss classification identifiers. After encoding the BIM components in the Revit models, further analysis was conducted on the relationship between EF and Ss classifications. As shown in Figure 6, the

correspondence between EF and Ss classifications in the dataset is not strictly one-to-one; rather, extensive one-to-many, many-to-one, and even many-to-many mapping structures are observed. In particular, for wall components, there is a prominent many-to-many relationship among EF_25_10_25 (External walls), EF_25_10_40 (Internal walls) and Ss_25_10_32_70 (Reinforced concrete wall systems), Ss_25_11_16 (Concrete wall systems), Ss_25_10_30_35 (Gypsum board partition systems), and Ss_25_12_60 (Panel enclosure systems). This indicates that special attention will be given to the classification performance of wall components in subsequent experiments. Such complex multi-level classification relationships reflect the functional and system-level multiplicity of building components, increasing the challenge of classification tasks. At the same time, it further validates the logical consistency and flexible extensibility of the Uniclass classification system at the Group and Subgroup hierarchy levels.

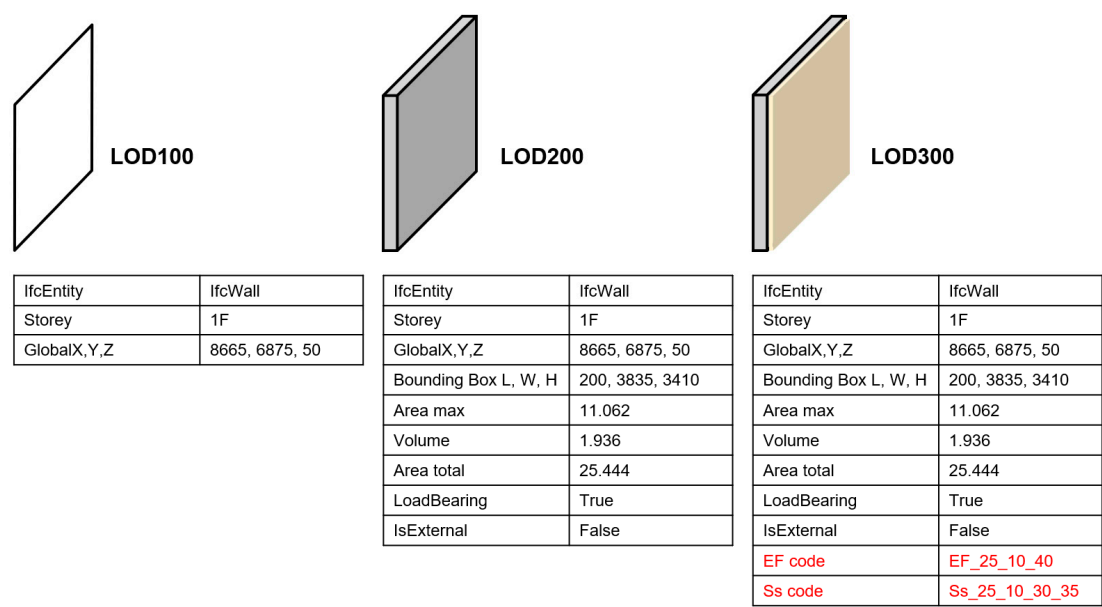


Figure 5. Correspondence between IFC Information and LOD Levels.

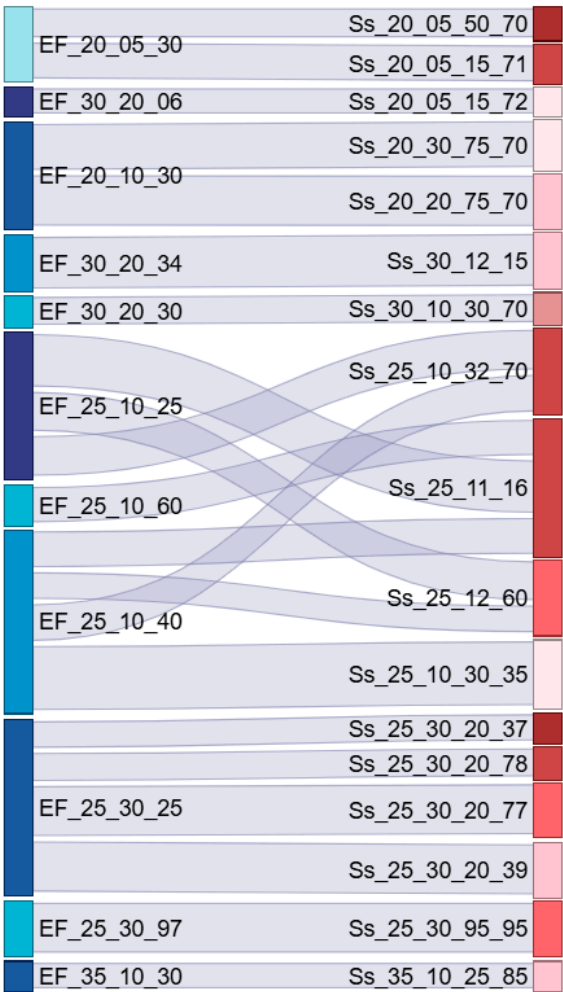


Figure 6. Mapping Relationships between EF and SS Classifications in the Dataset.

Figures 7 and 8 illustrate the distribution of category counts across the EF and Ss classification dimensions within the dataset. The results reveal an evident imbalance in the number of components among different categories for both EF and Ss classifications. Notably, wall-related components dominate across both classification dimensions, reflecting the ubiquity and critical importance of wall elements in real-world residential projects. Moreover, the EF and Ss classification systems provide relatively detailed categorization for wall components, implying that the classification accuracy for walls will significantly influence the overall performance of the model. In addition to walls, floor slabs and openings also constitute important portions of the dataset.

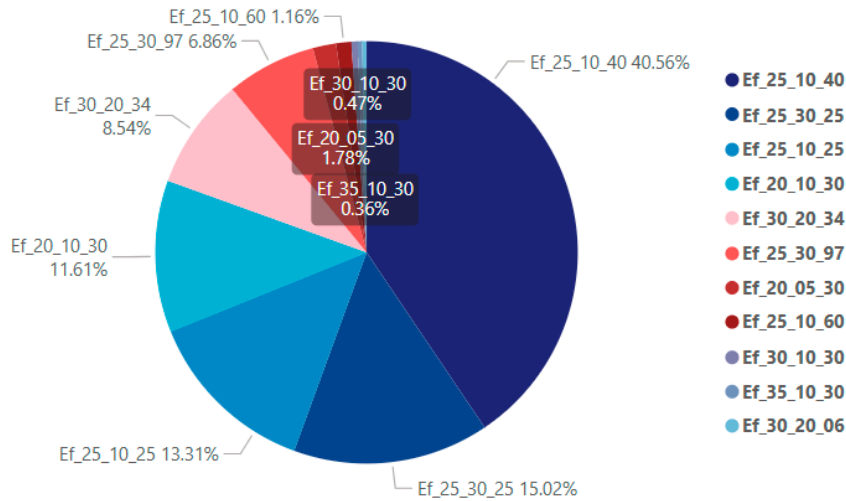


Figure 7. Composition Ratio of EF Codes in the Dataset.

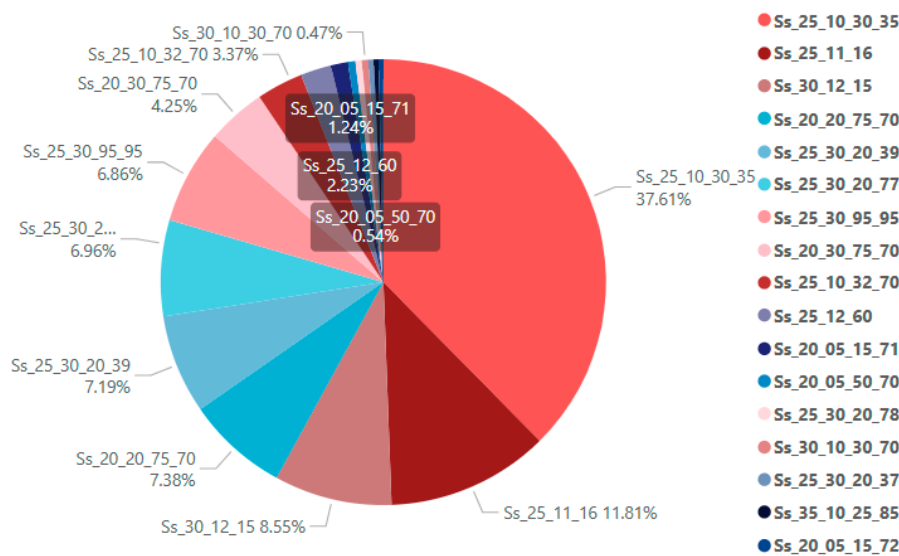


Figure 8. Composition Ratio of Ss Codes in the Dataset.

3.4. Evaluation Metrics

In this study, an exemplary IFC dataset was created and representative feature vectors were extracted to train machine learning classification models, thereby verifying the effectiveness of the proposed features for the automatic classification of building components. Following established practices from previous studies, the dataset was randomly split into a training set and a testing set at a ratio of 8:2 [20,31], containing 6,972 and 1,743 samples. During the training phase, 10-fold cross-validation was employed to enhance the generalization capability and robustness of the models.

For the evaluation of classification performance, four commonly used supervised learning metrics were adopted: Accuracy, Precision, Recall, and F1-score. These metrics were computed based on the confusion matrix shown in Table 3, and the specific calculation formulas are provided in Equations (1) to (4). Accuracy measures the proportion of correctly classified samples out of all test samples (Equation 1); however, in cases of class imbalance, accuracy alone may obscure the model's performance on minority classes and therefore should not be the sole evaluation metric. Recall focuses on the proportion of original positive samples that are correctly predicted, reflecting the model's coverage of the positive class (Equation 2); Precision measures the proportion of true positives among all samples predicted as positive (Equation 3); and the F1-score, being the harmonic

mean of Precision and Recall, provides a comprehensive measure of model performance under imbalanced data conditions (Equation 4).

Table 3. Truth Table Confusion Matrix.

	Actual Positive	Actual Negative
Predicted Positive	TP (True Positive)	FP (False Positive)
Predicted Negative	FN (False Negative)	TN (True Negative)

T: True (prediction is correct)
F: False (prediction is incorrect)
P: Positive (belongs to the target class)
N: Negative (does not belong to the target class)

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

In addition, to further analyze the contribution of each feature variable to the model's classification decisions, this study introduced a feature importance analysis. Within the adopted Random Forest algorithm, feature importance was quantified based on the Gini impurity criterion, which measures the degree to which each feature contributes to improving node purity during the decision-making process. Through this approach, key variables that have a significant impact on classification outcomes were identified. This analysis not only facilitates the optimization of feature selection strategies but also enhances the interpretability of the model, thereby supporting future improvements in component classification standardization and feature engineering efforts.

4. Experiments and Results

4.1. Data Preprocessing

Before training the model, it was necessary to standardize the original dataset. First, the non-numeric attributes “IsExternal” and “LoadBearing” were converted into one-hot encoded formats to meet the input requirements of machine learning models for numerical features. In addition, missing values were observed in the “Volume” of some IFC model components. To address this issue, K-Nearest Neighbors (KNN) imputation was applied for missing data filling. This method estimates missing values based on the distances between instances: for a sample with missing data, the five most similar samples (neighbors) in the dataset are selected, and the missing attribute is imputed using the average value of these neighbors. The implementation code is as follows:

```
KNN imputation code

volume_data = df[['Volume']]
imputer = KNNImputer(n_neighbors=5, weights='uniform')
volume_imputed = imputer.fit_transform(volume_data)
df['Volume'] = volume_imputed[:, 0]
```

4.2. Model Training

To achieve the automatic classification of BIM objects based on the EF and Ss codes within the Uniclass classification system, this study developed a supervised learning classification model using

the Random Forest algorithm. The model was implemented with the scikit-learn library, and the input feature vector consisted of 13 dimensions, covering semantic attributes (IfcEntity type, storey information, whether the object is external, and load-bearing status), spatial coordinates (GlobalX, GlobalY, GlobalZ, and bounding box dimensions), and dimensional information (maximum area, total area, and volume). The parameters of the Random Forest classifier were configured as follows: The model comprised 100 decision trees (n_estimators=100) to enhance classification stability and robustness. The Gini impurity (criterion="gini") was adopted as the splitting criterion to evaluate the purity gain after feature partitioning. No maximum depth (max_depth=None) was imposed, allowing each tree to grow fully according to the data characteristics. The minimum number of samples required to split an internal node was set to 2 (min_samples_split=2), and the minimum number of samples required at a leaf node was set to 1 (min_samples_leaf=1), ensuring the model's ability to capture fine-grained patterns even with small samples. Regarding feature selection, the number of features considered at each split was set to the default auto (equivalent to the square root of the total number of features), increasing diversity among trees and enhancing the overall ensemble performance. Bootstrap sampling (bootstrap=True) was enabled, meaning that each tree was trained using bootstrapped subsets of the data, further improving the model's generalization capability. To ensure reproducibility, the random seed was fixed at 42 (random_state=42). This configuration balances high classification performance with training efficiency and experimental reproducibility, making it suitable for the automatic classification of multi-category BIM objects in this study.

4.3. Experimental Results

To validate the classification performance of the constructed Random Forest (RF), this study conducted comparative experiments against Support Vector Machine (SVM) and Decision Tree (DT) algorithms. All three models were trained and tested on the same dataset, and their performance was evaluated on both EF and Ss coding classification tasks. The results were assessed using four metrics: Accuracy, Precision, Recall, and F1-score. Table 4 summarizes the classification performance of each model on the testing set.

The results indicate that the Random Forest model outperformed both the Decision Tree and SVM models in the EF coding task across all metrics, including Accuracy, Recall, and F1-score. Similarly, in the SS coding classification, Random Forest again demonstrated superior performance on all evaluation indices, showcasing stronger classification capability. Overall, the comparative evaluation confirmed that Random Forest is better suited than DT and SVM for the BIM object classification tasks addressed in this study. As an ensemble learning method composed of multiple decision trees, Random Forest exhibits excellent nonlinear modeling capabilities and strong noise resistance. It can maintain high classification accuracy even when handling high-dimensional features without requiring explicit feature selection, demonstrating superior generalization ability. Therefore, Random Forest presents significant advantages in the automatic classification of BIM objects.

Table 4. Classification results of different machine learning algorithms.

Model	Precision	Precision	Recall	Recall	F1-score	F1-score	Accuracy	Accuracy
	EF	Ss	EF	Ss	EF	Ss	EF	Ss
RF	0.99	0.99	0.99	0.94	0.99	0.96	1.00	0.99
DT	1.00	0.97	0.96	0.93	0.98	0.94	1.00	0.98
SVM	0.99	0.85	0.91	0.77	0.94	0.79	0.99	0.94

5. Discussion

5.1. Results Analysis

5.1.1. Classification Results Analysis

Based on a series of evaluation metrics and the analysis of the confusion matrices, it can be observed that the Random Forest algorithm demonstrated high performance in the component classification tasks, achieving an accuracy of 1.00 and 0.99 for the EF and SS classification tasks, respectively. Figure 9 illustrates the confusion matrix for the classification of 11 EF coding categories. Overall, the model exhibited strong classification capability across most categories, with the majority of samples correctly assigned to their corresponding classes. Specifically, the model performed exceptionally well in distinguishing EF_25_10_25 (External walls) and EF_25_10_40 (Internal walls), with 229 internal wall samples and 705 external wall samples correctly classified. This result indicates the model's ability to accurately differentiate between external and internal walls, with minimal misclassification.

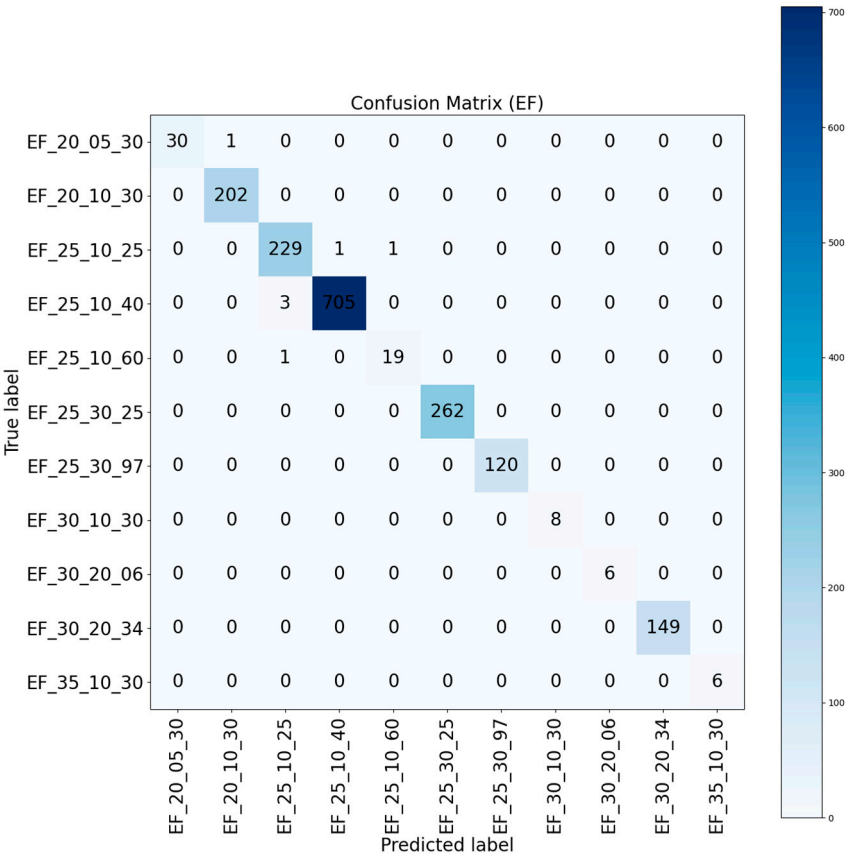


Figure 9. Confusion Matrix (EF).

However, the confusion matrix also revealed some misclassification among certain categories. In particular, EF_25_10_40 (Internal walls), EF_25_10_60 (Parapet walls), and EF_25_10_25 (External walls) exhibited cross-category misclassifications. For instance, one external wall sample was misclassified as an internal wall, one external wall sample was misclassified as a parapet wall, three internal wall samples were misclassified as external walls, and one parapet wall sample was misclassified as an external wall. Despite these four instances of confusion between internal and external walls, the overall classification accuracy remained high given the large sample size. Regarding the misclassification between parapet walls and external walls, further analysis revealed that both categories share the "IsExternal" attribute (True) and have considerable similarity in geometric features, making differentiation more challenging for the model. Additionally, the

relatively small sample size for parapet walls further contributed to lower classification accuracy in this category. Aside from the above confusions, misclassification among other categories was limited. The confusion matrix predominantly exhibited a dark diagonal pattern, with only a few darker off-diagonal cells, indicating robust overall classification performance. The model successfully achieved high recognition accuracy for most EF categories, demonstrating stable and reliable classification capability across different types of BIM objects. The distribution pattern in the confusion matrix further validates the effectiveness of the model in EF classification and also highlights potential directions for future refinement, particularly in distinguishing between highly similar categories.

Figure 10 presents the confusion matrix results for the Ss coding classification task, covering 17 system categories. Overall, the model demonstrated high classification accuracy across most system categories. Notably, the Ss_25_10_30_35 (Gypsum board partition systems) category achieved outstanding classification performance, with 654 correctly classified samples, making it the category with the highest sample count in the test set. This indicates that this system type had sufficient training samples and distinctive feature patterns conducive to accurate recognition. In contrast, the classification accuracy for Ss_25_10_32_70 (Reinforced concrete wall systems) and several door system categories was comparatively lower. For example, within the classification results for Ss_25_10_32_70 (Reinforced concrete wall systems), six samples were misclassified as Ss_25_11_16 (Concrete wall systems), indicating some confusion by the model in distinguishing different types of concrete wall systems. Additionally, a few misclassifications were observed for the Ss_25_12_60 (Panel enclosure systems), where two samples were incorrectly classified as Ss_25_10_30_35 (Gypsum board partition systems). This confusion is likely attributable to similarities in geometric dimensions or material properties between these system types, complicating the model's discrimination process.

Regarding door system classification, the accuracy for Ss_25_30_20_37 (High-security doorset systems) was relatively low, with some samples misclassified as Ss_25_30_20_39 (Hinged doorset systems). Similarly, samples from Ss_25_30_20_78 (Sliding folding doorset systems) were sometimes misclassified as Ss_25_30_20_39 (Hinged doorset systems) or Ss_25_30_20_77 (Sliding doorset systems). This can primarily be attributed to the high similarity in geometric features and opening/closing mechanisms among these door system categories, coupled with the relatively small sample size available for training, leading to less distinct decision boundaries.

In summary, the Random Forest model developed in this study exhibited robust performance in recognizing the system-level attributes of BIM objects, achieving high classification accuracy particularly for common categories such as foundations, beams, columns, and windows. However, a certain degree of confusion persists among system categories with structurally or functionally similar characteristics, suggesting that future research should focus on feature refinement or dataset expansion to further improve classification accuracy.

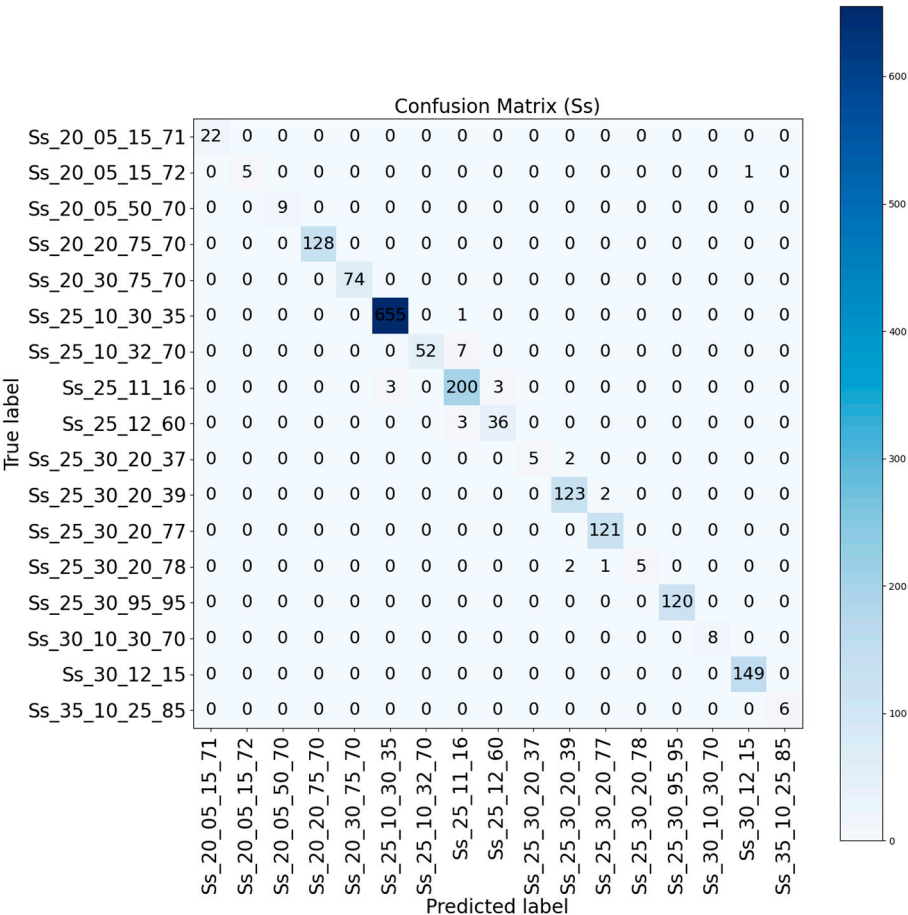


Figure 10. Confusion Matrix (Ss)

5.1.2. Feature Importance Analysis

Figures 11 illustrate the feature importance rankings of the Random Forest model used for the EF and Ss code classification tasks, respectively. In the EF coding classification task, the top three most important features were identified as IfcEntity, Bounding Box H, and IsExternal. Among these, IfcEntity represents the semantic category of the component, serving as the most distinctive structural feature, while Bounding Box H and IsExternal pertain to typical geometric and semantic boundary attributes. These results indicate that the model primarily relies on geometric dimensions and structural boundary characteristics—rather than purely spatial attributes—when distinguishing between functional categories such as walls, slabs, and foundations. Additionally, features like Volume and LoadBearing also contributed meaningfully to the classification task. It is noteworthy that global spatial information (such as GlobalX, GlobalY, GlobalZ) and the floor attribute (Storey) ranked lowest in feature importance, indicating a significantly lesser impact compared to geometric features. This suggests that the model's classification decisions are not based on the absolute spatial location or floor distribution of components but are instead driven by intrinsic geometric and structural properties. Consequently, the model demonstrates strong generalization capability and robustness. For the Ss coding classification task, the feature importance ranking showed a similar trend. IfcEntity and Bounding Box H again emerged as the top two features, while Bounding Box L, IsExternal, and Bounding Box W also exhibited considerable importance. This finding further confirms that, at the system classification level, the model primarily bases its decisions on volumetric and geometric attributes rather than global spatial coordinates.

In summary, the feature importance analysis indicates that the proposed classification model possesses strong independence from spatial distribution and achieves generalizable classification through geometry-driven learning. Such characteristics ensure the model's applicability even during

early design stages (Low LOD phases), laying a solid foundation for practical deployment of automatic BIM element classification in preliminary modeling workflows.

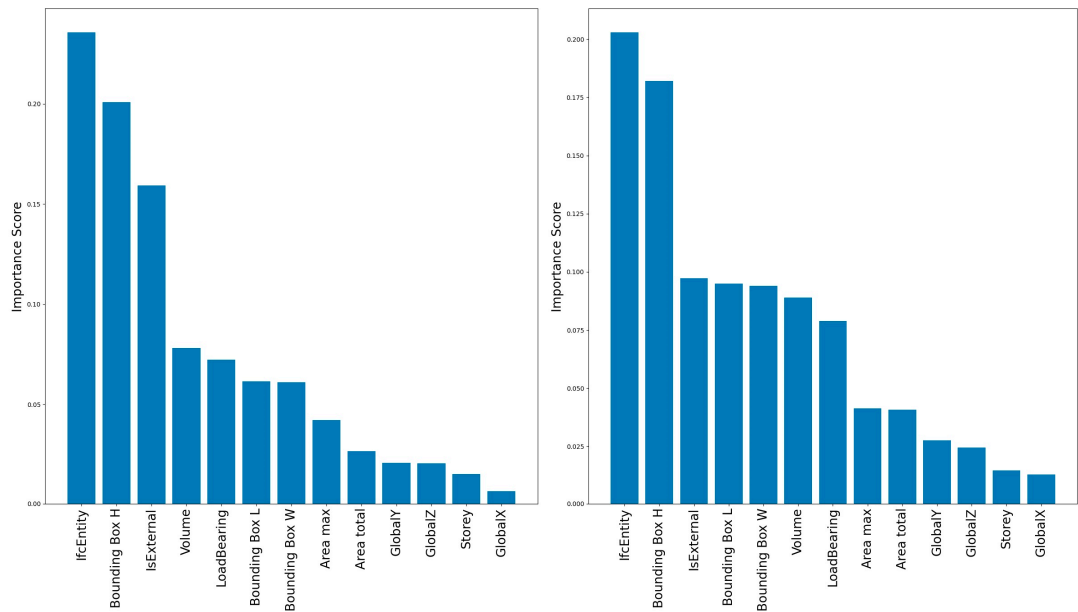


Figure 11. Feature Importance Analysis for Target Labels (Left: EF Labels, Right: SS Labels).

5.2. Contributions

Compared to existing research on the automatic classification of BIM objects, this study offers the following key contributions and innovations:

- A novel automatic classification method for BIM objects based on IFC data is proposed. By extracting feature variables from IFC data and training a Random Forest model, the method achieved over 99% classification accuracy in both EF and Ss coding tasks, demonstrating its effectiveness and robustness.
- The study successfully implements automatic classification and coding of BIM objects based on the Uniclass classification system (EF and Ss standards). This approach overcomes the traditional limitations of relying solely on geometric or functional features by employing a standardized classification framework, thereby enhancing the systematization, standardization, and engineering applicability of BIM object classification.
- For the first time, high-precision, fine-grained classification is achieved using only low-level detailed IFC data at LOD 200. The proposed method significantly reduces reliance on high-precision modeling data and substantially improves applicability during the early design stages (Early Design Stage) and low-LOD phases, offering a viable solution for intelligent management at the preliminary stage of BIM projects.

5.3. Comparison with Existing Methods

Currently, research focusing on the automatic classification of BIM objects based on standardized classification systems remains relatively limited. Most existing studies primarily adopt semantic analysis or image-based datasets to classify BIM objects. Despite differences in dataset types and specific methods, these studies, like the present work, confirm the feasibility of utilizing machine learning techniques for automatic BIM object classification [12,18,20].

For instance, Liu et al. demonstrated an approach that combines IFC data, image features, and semantic labels to automatically classify building components, achieving a classification accuracy exceeding 98%. However, their study mainly focused on the classification of object geometries, relying on complex data features that require significant data collection and preprocessing efforts

[11]. Similarly, Xu et al. proposed an automatic BIM object classification method based on IFC data targeting the OmniClass classification system, achieving 98.9% accuracy within the domain of prefabricated wall components. Nevertheless, their study was limited to prefabricated components, constraining its applicability, and primarily emphasized symbolic tagging of features rather than broader category classification [13].

In contrast, the method proposed in this study—an automatic classification approach based on the Uniclass system using low-LOD IFC data—demonstrates clear advantages in terms of data requirements, applicability, and systematization. The proposed method enables high-precision, multi-level component classification under conditions of low-detail modeling, showcasing superior practicality and engineering applicability, particularly for intelligent modeling assistance and information management during the early design stages.

5.4. Limitations and Future Work

Although this study successfully achieved automatic classification and coding of BIM elements based on IFC data and attained high classification accuracy in experiments, several limitations remain. The data used in this study were sourced exclusively from residential building projects, leading to certain limitations in the diversity of components and functional types. When applying the model to other types of buildings—such as commercial structures, industrial facilities, or infrastructure projects—new component categories not covered in the dataset may emerge, potentially resulting in a decline in classification accuracy. Future work should aim to expand the dataset to cover a broader range of building types, thereby enhancing the model's generalizability and adaptability. In addition, the classification labels for BIM objects in the dataset were manually assigned, making some degree of human error inevitable. Such annotation errors may affect the quality of supervisory signals during model training and consequently introduce noise into the final classification results. Future research could incorporate stricter annotation review mechanisms or adopt semi-automated labeling tools to improve the consistency and accuracy of the dataset annotations.

Future research will also focus on extending the application of BIM object classification to finer-grained levels within the Uniclass system, particularly by developing automatic classification methods based on the Products (Pr) table. Building upon the existing classification of EF (Element/Function) and Ss (Systems) layers, the integration of Pr (Products) layer information will enable a multi-dimensional linkage across functions, systems, and specific products. Through this expansion, it will become feasible to establish mappings between Pr codes and component unit costs, thereby enabling the prediction of product prices based on low-LOD BIM data. The ultimate objective is to facilitate precise control and optimization of component costs during the early design stage of building projects, providing data-driven support for economic analysis and decision-making in the construction process.

6. Conclusions

Fine-grained classification of BIM elements is essential for enhancing information interoperability and supporting BIM-based applications across the building lifecycle. This study proposes a method that enables automatic fine-grained classification and coding of BIM elements based on the Uniclass classification system, utilizing low Level of Development (LOD 200) IFC data. Specifically, semantic features, spatial features, and dimensional features were extracted from the IFC models to create feature vectors, which were then mapped to Element Function (EF) and Systems (Ss) codes within the Uniclass classification system using a Random Forest classification model. The proposed method significantly reduces reliance on high-LOD modeling data and achieved classification accuracies exceeding 99% in both EF and Ss tasks. The classification results demonstrated strong robustness, achieving high precision across multiple categories, thereby validating the effectiveness of the proposed feature framework and classification model in automatic BIM element recognition. This study further confirms the feasibility and application advantages of achieving efficient and scalable BIM element classification using standardized low-LOD BIM data.

Employing the standardized Uniclass classification system not only enhances the systematicity and standardization of classification results but also strengthens the interoperability and engineering applicability of the classification framework. Despite the positive outcomes, certain limitations remain. The current dataset is primarily sourced from residential building projects; applying the model to other building types may introduce variations in component types, potentially affecting the model's applicability and classification accuracy. Additionally, the classification labels in the dataset were manually assigned, introducing a degree of human error risk. Future research will aim to expand the dataset to include a wider range of building types and improve the consistency and accuracy of annotations. Moreover, further exploration will be conducted into incorporating the Products (Pr) table within the Uniclass classification system, combining EF and SS classification information to establish the linkage and prediction of component-level cost data. The ultimate objective is to enable accurate cost estimation and control during the early design stage using low-LOD BIM models, thereby extending the value of BIM in early-stage design applications and supporting intelligent decision-making capabilities.

Author Contributions: Conceptualization, S.T. and K.S.; methodology, S.T.; software, S.T.; validation, S.T.; investigation, S.T. and T.B.; writing—Original draft preparation, S.T.; writing—Review and editing, T.B. and K.S.; visualization, S.T.; funding acquisition, T.B. and K.S. All authors have read and agreed to the published version of the manuscript.

Funding: “This research was funded by Research grant of Construction Research Institute, grant number 2024-1” and “The APC was funded by Research grant of Construction Research Institute”. <https://www.kensetubukka.or.jp/trendtopics/subsidy/>.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT-4o for the purpose of polishing English expressions. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. EF and Ss Codes Included in the Dataset Used in This Study

Code	Title	Gr	Su	Se	Ob	SO	Description
EF_20_05_30	05EF	20	05	30			Foundations
EF_20_10_30	05EF	20	10	30			Framed structures
EF_25_10_25	05EF	25	10	25			External walls
EF_25_10_40	05EF	25	10	40			Internal walls
EF_25_10_60	05EF	25	10	60			Parapet walls
EF_25_30_25	05EF	25	30	25			Doors
EF_25_30_97	05EF	25	30	97			Windows
EF_30_10_30	05EF	30	10	30			Flat roofs
EF_30_20_06	05EF	30	20	06			Basement floors
EF_30_20_34	05EF	30	20	34			Ground floors
EF_35_10_30	05EF	35	10	30			External stairs
Ss_20_05_15_71	06Ss	20	05	15	71		Reinforced concrete pilecap and ground beam foundation systems
Ss_20_05_15_72	06Ss	20	05	15	72		Reinforced concrete raft foundation systems
Ss_20_05_50_70	06Ss	20	05	50	70		Reinforced concrete foundation and plinth systems

Ss_20_20_75_70	06Ss	20	20	75	70	Reinforced concrete beam systems
Ss_20_30_75_70	06Ss	20	30	75	70	Reinforced concrete column systems
Ss_25_10_30_35	06Ss	25	10	30	35	Gypsum board partition systems
Ss_25_10_32_70	06Ss	25	10	32	70	Reinforced concrete wall structure systems
Ss_25_11_16	06Ss	25	11	16		Concrete wall systems
Ss_25_12_60	06Ss	25	12	60		Panel enclosure systems
Ss_25_30_20_37	06Ss	25	30	20	37	High-security doorset systems
Ss_25_30_20_39	06Ss	25	30	20	39	Hinged doorset systems
Ss_25_30_20_77	06Ss	25	30	20	77	Sliding doorset systems
Ss_25_30_20_78	06Ss	25	30	20	78	Sliding folding doorset systems
Ss_25_30_95_95	06Ss	25	30	95	95	Window systems
Ss_30_10_30_70	06Ss	30	10	30	70	Reinforced concrete roof framing systems
Ss_30_12_15	06Ss	30	12	15		Concrete plank floor systems
Ss_35_10_25_85	06Ss	35	10	25	85	Suspended external stair systems

Gr: Group; Su: Sub group; Se: Section; Ob: Object; SO: Sub Object.

References

1. Lee, D.-G.; Park, J.-Y.; Song, S.-H. BIM-Based Construction Information Management Framework for Site Information Management. *Advances in Civil Engineering* **2018**, *2018*, 5249548, doi:10.1155/2018/5249548.

2. Xu, X.; Ma, L.; Ding, L. A Framework for BIM-Enabled Life-Cycle Information Management of Construction Project. *International Journal of Advanced Robotic Systems* **2014**, *11*, 126, doi:10.5772/58445.

3. BuildingSMART IFC. Available online: http://standards.buildingsmart.org/IFC/RELEASE/IFC4_1/FINAL/HTML/ (accessed on 8 September 2023).

4. Zhou, D.; Pei, B.; Li, X.; Jiang, D.; Wen, L. Innovative BIM Technology Application in the Construction Management of Highway. *Scientific Reports* **2024**, *14*, 15298, doi:10.1038/s41598-024-66232-5.

5. Ma, L.; Sacks, R.; Kattel, U.; Bloch, T. 3D Object Classification Using Geometric Features and Pairwise Relationships. *Computer-Aided Civil and Infrastructure Engineering* **2018**, *33*, 152–164, doi:10.1111/mice.12336.

6. Uniclass by NBS. Available online: <https://uniclass.thenbs.com/> (accessed on 7 May 2025).

7. Pupekis, D.; Navickas, A.A.; Klumbyte, E.; Seduikyte, L. Comparative Study of Construction Information Classification Systems: CCI versus Uniclass 2015. *Buildings* **2022**, *12*, doi:10.3390/buildings12050656.

8. Bloch, T.; Sacks, R. Comparing Machine Learning and Rule-Based Inferencing for Semantic Enrichment of BIM Models. *Automation in Construction* **2018**, *91*, 256–272, doi:10.1016/j.autcon.2018.03.018.

9. L. Choi; H. Kim; S. Kim; M. H. Kim Scalable Packet Classification Through Rulebase Partitioning Using the Maximum Entropy Hashing. *IEEE/ACM Transactions on Networking* **2009**, *17*, 1926–1935, doi:10.1109/TNET.2009.2018618.

10. Lim, Y.T.; Yi, W.; Wang, H. Application of Machine Learning in Construction Productivity at Activity Level: A Critical Review. *Applied Sciences* **2024**, *14*, doi:10.3390/app142210605.

11. Liu, H.; Gan, V.J.L.; Cheng, J.C.P.; Zhou, S. (Alexander) Automatic Fine-Grained BIM Element Classification Using Multi-Modal Deep Learning (MMDL). *Advanced Engineering Informatics* **2024**, *61*, 102458, doi:10.1016/j.aei.2024.102458.

12. M. Seydgar; É. A. Poirier; A. Motamedi Comparative Evaluation of Deep Neural Network Performance for Point Cloud-Based IFC Object Classification. *IEEE Access* **2024**, *12*, 108303–108312, doi:10.1109/ACCESS.2024.3436681.

13. Xu, Z.; Xie, Z.; Wang, X.; Niu, M. Automatic Classification and Coding of Prefabricated Components Using IFC and the Random Forest Algorithm. *Buildings* **2022**, *12*, doi:10.3390/buildings12050688.

14. Li, H.; Zhang, J.; Chang, S.; Sparkling, A. BIM-Based Object Mapping Using Invariant Signatures of AEC Objects. *Automation in Construction* **2023**, *145*, 104616, doi:10.1016/j.autcon.2022.104616.

15. Shan, Z.; Fu, D.; Qiu, L.; Liang, Y.; Huang, C. A Semi-Explicit Practical Coding Method for Prefabricated Building Component Parts in China. *Buildings* **2023**, *13*, doi:10.3390/buildings13051236.

16. Parece, S.; Resende, R.; Rato, V. A BIM-Based Tool for Embodied Carbon Assessment Using a Construction Classification System. *Developments in the Built Environment* **2024**, *19*, 100467, doi:[10.1016/j.dibe.2024.100467](https://doi.org/10.1016/j.dibe.2024.100467).
17. Jung, N.; Lee, G. Automated Classification of Building Information Modeling (BIM) Case Studies by BIM Use Based on Natural Language Processing (NLP) and Unsupervised Learning. *Advanced Engineering Informatics* **2019**, *41*, 100917, doi:[10.1016/j.aei.2019.04.007](https://doi.org/10.1016/j.aei.2019.04.007).
18. Utkucu, D.; Ying, H.; Wang, Z.; Sacks, R. Classification of Architectural and MEP BIM Objects for Building Performance Evaluation. *Advanced Engineering Informatics* **2024**, *61*, 102503, doi:[10.1016/j.aei.2024.102503](https://doi.org/10.1016/j.aei.2024.102503).
19. Yu, Y.S.; Kim, S.H.; Lee, W.B.; Koo, B.S. Ensemble-Based Deep Learning Approach for Performance Improvement of BIM Element Classification. *KSCE Journal of Civil Engineering* **2023**, *27*, 1898–1915, doi:[10.1007/s12205-023-2331-y](https://doi.org/10.1007/s12205-023-2331-y).
20. Koo, B.; La, S.; Cho, N.-W.; Yu, Y. Using Support Vector Machines to Classify Building Elements for Checking the Semantic Integrity of Building Information Models. *Automation in Construction* **2019**, *98*, 183–194, doi:[10.1016/j.autcon.2018.11.015](https://doi.org/10.1016/j.autcon.2018.11.015).
21. Austern, G.; Bloch, T.; Abulafia, Y. Incorporating Context into BIM-Derived Data—Leveraging Graph Neural Networks for Building Element Classification. *Buildings* **2024**, *14*, doi:[10.3390/buildings14020527](https://doi.org/10.3390/buildings14020527).
22. Akinosho, T.D.; Oyedele, L.O.; Bilal, M.; Ajayi, A.O.; Delgado, M.D.; Akinade, O.O.; Ahmed, A.A. Deep Learning in the Construction Industry: A Review of Present Status and Future Innovations. *Journal of Building Engineering* **2020**, *32*, 101827, doi:[10.1016/j.jobe.2020.101827](https://doi.org/10.1016/j.jobe.2020.101827).
23. Dopazo, D.A.; Mahdjoubi, L.; Gething, B.; Mahamadu, A.-M. An Automated Machine Learning Approach for Classifying Infrastructure Cost Data. *Computer-Aided Civil and Infrastructure Engineering* **2024**, *39*, 1061–1076, doi:<https://doi.org/10.1111/mice.13114>.
24. Antoniou, F.; Aretoulis, G.; Giannoulakis, D.; Konstantinidis, D. Cost and Material Quantities Prediction Models for the Construction of Underground Metro Stations. *Buildings* **2023**, *13*, doi:[10.3390/buildings13020382](https://doi.org/10.3390/buildings13020382).
25. Jianzu, W.; Chaojie, Z. Machine Learning in Strategic Management Research: A Review and Prospects. *Foreign Economics & Management* **2025**, *47*, 119–136, doi:[10.16538/j.cnki.fem.20240409.101](https://doi.org/10.16538/j.cnki.fem.20240409.101).
26. Kim, M.-K.; Thedja, J.P.P.; Chi, H.-L.; Lee, D.-E. Automated Rebar Diameter Classification Using Point Cloud Data Based Machine Learning. *Automation in Construction* **2021**, *122*, 103476, doi:[10.1016/j.autcon.2020.103476](https://doi.org/10.1016/j.autcon.2020.103476).
27. Zeng, Z.; Ying, G.; Zhang, Y.; Gong, Y.; Mei, Y.; Li, X.; Sun, H.; Li, B.; Ma, J.; Li, S. Classification of Failure Modes, Bearing Capacity, and Effective Stiffness Prediction for Corroded RC Columns Using Machine Learning Algorithm. *Journal of Building Engineering* **2025**, *102*, 111982, doi:[10.1016/j.jobe.2025.111982](https://doi.org/10.1016/j.jobe.2025.111982).
28. Li, Y.; Ouyang, W.; Xin, Z.; Zhang, H.; Sun, S.; Zhang, D.; Zhang, W. Machine Learning for Defect Condition Rating of Wall Wooden Columns in Ancient Buildings. *Case Studies in Construction Materials* **2025**, *22*, e04458, doi:[10.1016/j.cscm.2025.e04458](https://doi.org/10.1016/j.cscm.2025.e04458).
29. Mitera-Kielbasa, E.; Zima, K. Automated Classification of Exchange Information Requirements for Construction Projects Using Word2Vec and SVM. *Infrastructures* **2024**, *9*, doi:[10.3390/infrastructures9110194](https://doi.org/10.3390/infrastructures9110194).
30. Romero-Jarén, R.; Arranz, J.J. Automatic Segmentation and Classification of BIM Elements from Point Clouds. *Automation in Construction* **2021**, *124*, 103576, doi:[10.1016/j.autcon.2021.103576](https://doi.org/10.1016/j.autcon.2021.103576).
31. KIEU, T.C.; 志手一哉 ファセット型分類体系を用いた WBS の構築におけるテーブル間連携の可能性に関する分析. 日本建築学会技術報告集 **2022**, *28*, 986–991, doi:[10.3130/aijt.28.986](https://doi.org/10.3130/aijt.28.986).
32. Emunds, C.; Pauen, N.; Richter, V.; Frisch, J.; van Treeck, C. SpaRSE-BIM: Classification of IFC-Based Geometry via Sparse Convolutional Neural Networks. *Advanced Engineering Informatics* **2022**, *53*, 101641, doi:[10.1016/j.aei.2022.101641](https://doi.org/10.1016/j.aei.2022.101641).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.